RESOURCES MANAGEMENT IN MULTI-CHANNEL RELAYING

by

Mahdi Hajiaghayi

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto

Copyright © 2012 by Mahdi Hajiaghayi
Abstract

Resource Management in Multi-channel Relaying

Mahdi Hajiaghayi
Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto
2012

Resource management—particularly power and spectrum management—is becoming increasingly important owing to the fast growing market of smart-phones and other power-hungry wireless devices. While WiFi and cellular communication accounts for a significant portion of the smart-phone power expenditure, spectrum is equally paramount and scarce as well. This demands for an efficient and judicious resource management schemes that is also viable in terms of the practical implementation and complexity. This thesis focuses on the various setups of multi-channel relaying system as an emerging wireless technology, and provides rate optimal, yet easy-to-implement, resource management solutions for them. We exploit the channel pairing (CP) capability of a multi-channel relay node in our design. This capability allows the relay to receive a signal from one channel and transmit a processed version of the signal on a different channel. CP jointly optimized with power allocation (PA), which determines each channel’s power, can lead to significant improvement in spectral efficiency. For two setups, namely multi-hop and multi-user setups, we present the total achievable rates through optimizing CP, PA and channel-user assignment which incurs multi-user diversity. While the achievable rates provide theoretical insight for the performance of such systems, we next incorporate the integer nature of bit loading and rate adaptation, and via an innovative optimization technique, we present the jointly optimal solution to the problem of bit loading, PA and CP.
Acknowledgements

This thesis was part of a five year research carried out in wireless computing lab (WCL) at the University of Toronto. During my Ph.D program, I met many exceptional individuals whom I would like to mention here.

First, I owe my deepest gratitude to my supervisors, Prof. Liang and Prof. Dong for their endless support and guidance throughout my studies. Our constructive meetings offered opportunities to refine my ideas and to draw inspiration in completing our work. I also thank them for providing me with the freedom and flexibility to develop my research project.

Next, I would like to thank my proposal and defense committee, Prof. Yu and Prof. Adve and the external examiner, Prof. Jiang, for their valuable time reading my thesis and for sharing their fruitful comments and feedbacks.

My colleagues at the wireless computing lab: Dr. Lotfinejad, Dr. Farbod and fellow students Hossein, Amir, Rostom, Lei, Guang, facilitated my research with their valuable hints and extensive discussions. Without their help, completing this thesis would have been very difficult.

Since my arrival in Toronto, I have had an amazing and adventurous life in a unique house shared with great and friendly people from all corners of the globe. Through them, I have gained a genuine Canadian experience as well as precious international knowledge. My special thanks goes to Larissa Satta for her passion for conversing about many important matters of life as well as occasionally revising my technical wrtiting. Furthermore, I would like to acknowledge Dr. Emmanuel Bongajum, Jamen Dunnings, Evan Karam, Anna Pacchioni, Anthony Durand and many more for their willingness to share their bright thoughts and experiences which have shaped my life path and provided me with a great living experience. I was also privileged to have my long-time friend, Dr. Keyhan Kobravi, nearby in Toronto, a person whose company is always a delight.

I am very grateful to Dr. Wijting and Dr. Berg at the Nokia research center for giving me the opportunity to work there as a research intern. During the 5-month internship program, I much
enjoyed collaborating and interacting with my mindful and kind colleagues particularly Dr. Mikka Rinne and Dr. Cassio Ribeiro. This enabled me to enrich my theoretical and practical knowledge of wireless communication. Dr. Carl Wijting’s mentorship benefited me greatly through his generous help and sharing of ideas concerning work as well as other interesting aspects of life.

Last but certainly not least, my deepest gratitude goes to my parents and my brother, Prof. M.T Hajiaghayi, for their support and help through the entire course of my life. I’d also like to thank my adorable sisters, Monir and Mehri, for their encouragement and loving support. Despite the fact that we live far apart, their warm and endless support was always with me and in many ways enabled me to complete my Ph.D program.
Contents

1 Introduction .............................................. 1
   1.1 Thesis Contributions ................................. 3

2 Background and Related Work ............................ 7
   2.1 Multi-channel Relaying System ......................... 7
   2.2 Optimization Tools and Techniques ...................... 18
   2.3 Related Works ......................................... 26

3 Jointly Optimal CP and PA for Multi-channel Multi-hop Relaying ........................................... 35
   3.1 System Model and Problem Statement ...................... 36
   3.2 Optimal Multi-hop CP Under Fixed PA ..................... 41
   3.3 Jointly Optimal CP and PA: A Separation Principle ........... 44
   3.4 Numerical Results ........................................ 55
   3.5 Summary ................................................... 64

4 Jointly Optimal Channel Assignment and PA for Dual-Hop Multi-channel Multi-user Relaying ..................... 65
   4.1 System Model and Problem Statement ...................... 66
   4.2 Weighted Sum-Rate Maximization for Multi-channel DF ........... 69
   4.3 Extensions to General Relaying Strategies .................. 82
   4.4 Numerical Results ........................................ 83
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5 Summary</td>
<td>90</td>
</tr>
<tr>
<td>5 Jointly Optimal Bit Loading, CP and PA for Multi-channel Relaying</td>
<td>91</td>
</tr>
<tr>
<td>5.1 System Model and Problem Statement</td>
<td>92</td>
</tr>
<tr>
<td>5.2 Upper Bound to Optimal Bit Rate via Dual Decomposition</td>
<td>96</td>
</tr>
<tr>
<td>5.3 Extraction of Jointly Optimal Solution</td>
<td>98</td>
</tr>
<tr>
<td>5.4 Complexity Reduction and Analysis</td>
<td>104</td>
</tr>
<tr>
<td>5.5 Numerical Results</td>
<td>108</td>
</tr>
<tr>
<td>5.6 Summary</td>
<td>109</td>
</tr>
<tr>
<td>6 Conclusion</td>
<td>111</td>
</tr>
<tr>
<td>7 Appendices</td>
<td>113</td>
</tr>
<tr>
<td>7.1 Appendix A</td>
<td>113</td>
</tr>
<tr>
<td>7.2 Appendix B</td>
<td>117</td>
</tr>
<tr>
<td>7.3 Appendix C</td>
<td>124</td>
</tr>
<tr>
<td>References</td>
<td>126</td>
</tr>
</tbody>
</table>
List of Figures

2.1 Basic topology of a multi-channel relaying system .......................... 8
2.2 Source-destination pair in multi-channel (OFDM) system ................... 10
2.3 OFDM relay illustration ......................................................... 11
2.4 The rate performance of pairing, piping and direct communication for 8-channel DF relay system .......................................................... 17

3.1 Illustration of multi-channel multi-hop relaying network with channel pairing . 37
3.2 Three-hop relay with two channels ................................................. 42
3.3 Converting an (L+1)-hop relaying to an equivalent 3-hop relaying. (a) An (L+1)-hop relaying network; (b) An equivalent 3-hop relaying network. .... 45
3.4 Normalized rate vs. the average SNR for DF OFDM relaying with $M = 4$ and $N = 64$ under total power constraint ........................................... 58
3.5 Normalized rate vs. the average SNR for AF OFDM relaying with $M = 4$ and $N = 64$ under total power constraint ........................................... 58
3.6 Normalized rate vs. the average SNR for DF OFDM relaying with $M = 4$ and $N = 16$ under individual power constraint ......................................... 59
3.7 Normalized rate vs. the average SNR for AF OFDM relaying with $M = 4$ and $N = 16$ under individual power constraint ......................................... 59
3.8 Normalized rate vs. number of taps for DF OFDM relaying with $M = 4$, $N = 64$, and $\text{SNR}_{\text{avg}} = 12 \text{dB}$ under total power constraint ............... 61
3.9 Normalized rate vs. number of taps for AF OFDM relaying with $M = 4$, $N = 64$, and $\text{SNR}_{\text{avg}} = 12\text{dB}$ under total power constraint. ............................................. 61
3.10 Normalized rate vs. number of channels for DF OFDM relaying with $M = 4$ and $\text{SNR}_{\text{avg}} = 12\text{dB}$ under total power constraint. ............................................. 62
3.11 Normalized rate vs. number of channels for AF OFDM relaying with $M = 4$ and $\text{SNR}_{\text{avg}} = 12\text{dB}$ under total power constraint. ............................................. 62
3.12 Normalized rate vs. number of hops for DF OFDM relaying with $N = 64$ and $\text{SNR}_{\text{avg}} = 12\text{dB}$ under total power constraint. ............................................. 63
3.13 Normalized rate vs. number of hops for AF OFDM relaying with $N = 64$ and $\text{SNR}_{\text{avg}} = 12\text{dB}$ under total power constraint. ............................................. 64
4.1 Illustration of dual-hop multi-channel relaying. .................................................. 66
4.2 Simulation configuration with $K = 4$ users ......................................................... 85
4.3 Normalized weighted sum-rate vs. nominal SNR with $w = [.25, .25, .25, .25]$, $N = 16$, $K = 4$, and DF relaying. ................................................................. 85
4.4 Normalized weighted sum-rate vs. nominal SNR with $w = [.15, .15, .35, .35]$, $N = 16$, $K = 4$, and DF relaying. ................................................................. 86
4.5 Normalized weighted sum-rate vs. number of users with equal weight $w_i = 1$, for $1 \leq i \leq K$, $N = 16$, and DF relaying. ................................................................. 87
4.6 Number of iterations vs. number of channels for DF OFDMA relaying with $K = 4$, equal non-equal weight under total power constraint .............................. 88
4.7 Normalized weighted sum-rate vs. relay location; $K = 4$, $w = [.15, .15, .35, .35]$, $N = 16$, and DF relaying. ................................................................. 89
4.8 Normalized weighted sum-rate vs. relay location; $K = 4$, $w = [.15, .15, .35, .35]$, $N = 16$, and AF relaying. ................................................................. 89
5.1 Function $B^*(P_t)$. The point $(P^*(\lambda), B^*(\lambda))$ is on a corner, and $g(\lambda)$ passes through it. ................................................................. 100
5.2 Function $P^*(\lambda)$. It is discontinuous at $\lambda = \lambda^*$. ............................. 102

5.3 Per channel rate vs. $\text{SNR}_{avg}^{R-D}$ for multi-channel AF relaying with $\text{SNR}_{avg}^{S-R} = 10dB$. ................................................................. 110

5.4 Per channel rate vs. number of channels for multi-channel AF relaying with $\text{SNR}_{avg}^{S-R} = \text{SNR}_{avg}^{R-D} = 10dB$. .................................................. 110
# List of Notations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>min((a, b))</td>
<td>Minimum value of (a) and (b)</td>
</tr>
<tr>
<td>(I_N)</td>
<td>(N \times N) Identity matrix</td>
</tr>
<tr>
<td>((\cdot)^T)</td>
<td>Matrix transpose</td>
</tr>
<tr>
<td>((\cdot)^H)</td>
<td>Conjugate transpose of the matrix argument</td>
</tr>
<tr>
<td>\text{diag}[..]</td>
<td>Diagonal matrix operator</td>
</tr>
<tr>
<td>(\mathbb{E}{\cdot})</td>
<td>Expectation operation</td>
</tr>
<tr>
<td>(A)</td>
<td>The Matrix (A)</td>
</tr>
<tr>
<td>(a)</td>
<td>The vector (a)</td>
</tr>
<tr>
<td>(\text{det}{A})</td>
<td>Determinant of the matrix (A)</td>
</tr>
<tr>
<td>(|A|)</td>
<td>Frobenius norm of the matrix (A)</td>
</tr>
<tr>
<td>(CN(\mu, \sigma^2))</td>
<td>A circularly symmetric complex Gaussian random variable (RV) with mean (\mu) and variance (\sigma^2)</td>
</tr>
<tr>
<td>(A \succeq 0)</td>
<td>Matrix (A) is positive semidefinite</td>
</tr>
<tr>
<td>(\Delta^2(f))</td>
<td>Hessian: the square matrix of second-order partial derivatives of (f)</td>
</tr>
</tbody>
</table>
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OFDM</td>
<td>orthogonal frequency division multiplexing</td>
</tr>
<tr>
<td>OFDMA</td>
<td>orthogonal frequency division multiplexing access</td>
</tr>
<tr>
<td>CP</td>
<td>channel pairing</td>
</tr>
<tr>
<td>PA</td>
<td>power allocation</td>
</tr>
<tr>
<td>AF</td>
<td>amplify-and-forward</td>
</tr>
<tr>
<td>DF</td>
<td>decode-and-forward</td>
</tr>
<tr>
<td>CF</td>
<td>compress-and-forward</td>
</tr>
<tr>
<td>NP</td>
<td>non-deterministic polynomial-time</td>
</tr>
<tr>
<td>FFT</td>
<td>fast fourier transform</td>
</tr>
<tr>
<td>IFFT</td>
<td>inverse fast fourier transform</td>
</tr>
<tr>
<td>PSK</td>
<td>pulse shift keying</td>
</tr>
<tr>
<td>QAM</td>
<td>quadrature amplitude modulation</td>
</tr>
<tr>
<td>QPSK</td>
<td>quadrature pulse shift keying</td>
</tr>
<tr>
<td>BPSK</td>
<td>binary pulse shift keying</td>
</tr>
<tr>
<td>PAM</td>
<td>pulse amplitude modulation</td>
</tr>
<tr>
<td>ISI</td>
<td>inter-symbol interference</td>
</tr>
<tr>
<td>ADC</td>
<td>analog to digital converter</td>
</tr>
<tr>
<td>DAC</td>
<td>digital to analog converter</td>
</tr>
<tr>
<td>AWGN</td>
<td>additive white Gaussian noise</td>
</tr>
<tr>
<td>P/S</td>
<td>parallel to serial</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>S/P</td>
<td>serial to parallel</td>
</tr>
<tr>
<td>IP</td>
<td>integer programming</td>
</tr>
<tr>
<td>AP</td>
<td>assignment problem</td>
</tr>
<tr>
<td>SNR</td>
<td>signal to noise ratio</td>
</tr>
<tr>
<td>DCDM</td>
<td>divide-and-conquer dual minimization</td>
</tr>
<tr>
<td>RV</td>
<td>random variable</td>
</tr>
<tr>
<td>MRC</td>
<td>maximum ratio combining</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

The emerging next-generation wireless systems adopt a multi-channel relaying architecture for broadband access and coverage improvement [1][2]. The superiority of a multi-channel system over a single-channel narrowband system stems from its attractive implementation features such as simpler equalizer and easy-to-build transceivers as well as its performance features such as higher spectral efficiency, reduced sensitivity to delay and time synchronization, resistance to fading and robustness against frequency selective communication channels [3]. Relaying, on the other hand, in comparison with the traditional direct communication offers potential diversity increase and better quality of service [4][5].

It appears judicious and adaptive resource allocation is crucial to fully realize the benefits of multi-channel relaying [6][7][8][9]. Resource allocation in such system involves more technical challenges as well as more design freedom compared with the traditional single-channel system. Most notably, as opposed to a narrow-band single-channel relay, a multi-channel relay has an additional frequency dimension available. It may receive a signal from one channel and transmit a processed version of the signal on a different channel. This multi-channel relaying capability can be exploited to choose the forwarding channel adaptively for the incoming signals, taking advantage of the diverse strength of different channels. Without such consideration, directly applying a relaying method optimized in single-carrier systems will be suboptimal for
a multi-channel relaying system.

The general problem of resource allocation for various multi-channel relaying scenarios considered in this thesis involves: 1) channel pairing: the pairing of incoming and outgoing channels at the relay; 2) power allocation: the determination of power used to transmit signals on these channels 3) channel-user assignment: assigning a subset of incoming-outgoing subcarrier pairs to a user for the case of multi-users orthogonal frequency division multiplexing access (OFDMA) relaying network. Since the channel condition on the same channel can vary drastically for different users, optimal channel pairing and channel-user assignment can potentially lead to significant improvement in spectral efficiency. Lastly, appropriate power allocation across channels at the source and the relay is essential to maximize the cooperative relaying performance.

Joint resource allocation, despite providing more design freedom, also introduces a new set of design challenges and complexity. One main challenge confronting the optimal resource allocation is that the problem formulation normally falls in the class of mixed integer programming problems whose solutions often bear prohibitive computational complexity. This is why the previous attempts to optimize this problem often either considered only a subset of the three problems above or adopt simpler degraded system models [10], [11], [12], [13], [14]. Several suboptimal approaches providing reduced complexity were also proposed under certain system models [11], [14], [13], [15].

So far, the problem was to optimally determine channel pairing and allocate power among different channels within a given power budget so as to maximize the information theoretic data rate. However, the proposed allocation schemes lead to non-integer rate adaption. Although they provide theoretical understanding for the performance of such systems, the solutions obtained are not suited for practical systems, where finite modulation formats and integer-valued bits are used for data transmission. With the practical constraint of discrete modulation level, the power allocation problem becomes even more challenging: how can the power and bits be allocated optimally to each channel, in order to maximize the system performance? In the
literature of point-to-point communications, this problem is generally referred to as bit loading design. Relaying adds another dimension to the problem, since the incoming and outgoing channels at the relay need to be optimally paired for maximal performance. In particular, the bit allocation on an incoming channel of the relay is no longer the same as that of the same channel on the outgoing side.

1.1 Thesis Contributions

In this thesis we explore various multi-channel relaying setups for which we obtain a viable optimal resource allocation solution to maximize the overall system performance. After reviewing the basic structure of a multi-channel relaying system followed by essential optimization tools and techniques in Chapter 2, we focus on the three main topics as follows:

Multi-channel Multi-hop Relaying [16] [17]: In Chapter 3, we first study the problem of resource allocation in multi-hop multi-channel relaying network to maximize the total achievable rate. We address both CP and PA issues at each relay which can be amplify-and-forward (AF) or decode-and-forward (DF). In general, for multi-hop relaying, there is strong correlation between CP and PA. Intuitively, to maximize the source-destination rate performance, the choice of CP at each relay would affect the choices of CP at other relays, which further depends on the specific PA scheme used. The optimal system performance requires joint consideration of CP and PA.

We show that the joint problem can be decoupled into two subproblems solved separately: first CP optimization, and then PA optimization. The decoupling of CP and PA optimization significantly reduces the problem search space and reduces the complexity of optimal solution. It follows that the channel pairing problem in joint optimization can be again decomposed into independent pairing problems at each relay based on sorted channel gains. Specifically, The channels at two consecutive hops are optimally paired according to their channel gain order, without the need for knowledge of power allocation on each channel. The resulting
decomposition allows simple distributed relay implementation for optimal operation, as well as easily adapting to the network topology changes.

The solution for optimizing power allocation for DF relaying is also provided through a dual-decomposition approach with which we can characterize the interaction among the nodes for power determination on a multi-hop path. For AF relaying, an asymptotically optimal PA solution is provided. These results are further extended to the multi-destination scenarios.

**Multi-channel Multi-user Relaying** [18][19]: Next, in Chapter 4 we consider a multi-user scenario where we specifically tackle the problem of channel-user assignment beside CP and PA for multi-channel relaying communication. The problem typically arises in cellular communication or wireless local area networks, through either dedicated relay stations or users temporarily serving as relay nodes. With interactions between multiple channels and multiple users, judicious channel-user assignment, which results in multi-user diversity, and CP together with PA can potentially lead to significant enhancement in spectral efficiency. However, the combinatorial nature of channel pairing and assignment generally leads to a mixed-integer programming problem.

We show that there is an efficient method to jointly optimize channel pairing, channel-user assignment, and power allocation in such general dual-hop relaying networks. The proposed solution framework is built upon continuous relaxation and Lagrangian dual minimization. Although this approach is often applied to integer programming problems [20], it generally provides only heuristic or approximate solutions. However, by exploring the rich structure in our problem, we show that judicious reformulation and choices of the optimization trajectory can preserve both the binary constraints and the strong Lagrangian duality property of the continuous version (i.e., zero duality gap), thus enabling a jointly optimal solution.

Through reformulation, we transform the core of the original problem into a special incidence of the class of three-dimensional assignment problems, which is NP hard in general but has polynomial-time solutions – in terms of the number of channels and users – for our specific setting of channel pairing and channel-user assignment. For the often studied con-
C H A P T E R 1. I N T R O D U C T I O N

ventional DF relaying with a maximum weighted sum-rate objective, we further propose a divide-and-conquer algorithm for dual minimization, which guarantees that convergence to an optimal solution requires only a polynomial number of iterations in the number of channels. This ensures the scalability of the proposed solution to large multi-channel systems.

Our proposed solution is applicable to a wide range of system configurations, accommodating both total and individual power constraints, and allowing direct source-destination links in relaying. We show that it can be modified to work with various relaying strategies in addition to DF, including variants of compress-and-forward (CF) and amplify-and-forward (AF). It gives jointly optimal solutions for any relaying strategies whose achievable rate is a concave function in the transmission powers. Through simulation and numerical comparison, we further illustrate that there is often a large performance gap between the jointly optimal solution and the suboptimal alternatives. The availability of our jointly optimal solution also allows us to numerically study the effect of different factors on the performance of optimal and suboptimal strategies.

Optimal Bit Loading and Multi-channel Relaying [21]: In the last chapter, Chapter 5, we consider the bit loading problem along with PA and CP problems in a dual-hop multi-channel relaying system. We aim to derive a practical and computationally efficient resource allocation method that leads to optimal channel selection and power allocation for a multi-channel capable relaying network with integer bit loading. Three issues of CP, PA and bit-loading are jointly addressed for a multi-channel relaying setup to maximize the overall bit rate. The resulting optimization problem is of non-convex mixed-integer programming in nature, with both a discrete objective and a discrete feasible region. Although the Lagrange duality approach can be applied to such problems [22], it generally does not lead to an optimal solution due to the non-zero duality gap. Nonetheless, we develop our solution through the Lagrange dual approach. By exploring the structure of our problem, we are able to bound the gap to the original objective to be within one bit. This knowledge allows us to extract the exact optimal integer solution. We further develop numerical techniques to reduce the algorithm
runtime. Despite the general NP-hardness of other mixed-integer optimization problems, we demonstrate that the proposed algorithm has a computational complexity that is polynomial in the number of channels. This suggests that it is more amenable to practical implementation than combinatoric search approaches, especially for practical systems, such as IEEE 802.16 OFDMA, with hundreds to thousands of frequency sub-channels.
Chapter 2

Background and Related Work

In this chapter, we give an overview of the basic structure of a multi-channel relaying system, its system model and achievable rate. We then review the general optimization techniques that we will frequently use throughout this thesis. Related work and literature review are presented in the last section of this chapter.

2.1 Multi-channel Relaying System

Multi-channel or multi-carrier physical-layer technology is proposed for next-generation wireless systems [23]. Particularly, Orthogonal Frequency Division Multiplexing (OFDM) as a special case of multi-channel transmission is widely adopted by the emerging wireless standards such as LTE uplink/downlink and IEEE 802.16 due to its attractive implementation features. It divides a broadband frequency-selective channel to a series of flat fading channels on which a parallel block of data is transmitted. Due to the orthogonality of the channels, no band-guard is required to separate the carriers, thus making OFDM highly spectrally efficient. It also makes OFDM resistant to fading and robust against frequency selective communication channels as all subchannels are flat fading. This gives rise to a simple equalizer at the OFDM receiver. As will be described in the next section, OFDM can be easily implemented using FFT and IFFT blocks. Furthermore, because of narrower bandwidth and longer time dimension of each flat
fading channel, an OFDM system is less sensitive to delay and time synchronization.

Meanwhile, relaying is a proven technique for capacity and transmission range enhancement, where a faraway or deeply shadowed user can be served with the help of relay nodes (users) that have a more favorable connection with the base station.

Future system designs are therefore evolving toward the adoption of a combination of multi-channel transmission and relaying architecture to allow broadband access and coverage improvement [1],[2]. In the ensuing subsection, we describe the general architecture of a multi-channel relaying system.

### 2.1.1 System Architecture

The basic topology for a multi-channel relaying system, as depicted in Fig. 2.1, is a three-node dual-hop scenario where a source communicates with a destination with the assistance of a relay. We assume a broadband multi-channel system where the available communication bandwidth is partitioned into an equivalent set of $N$ channels, each being used for sending a block of information symbols. A transmission from the source to the destination takes place in two phases. In the first phase, the source transmits a data block to the relay and destination on each channel. In the second phase, the relay forwards a processed version of the received signals over the outgoing channels. Below, we will describe the underlying structure of source-destination pair and the relay in further detail.

![Figure 2.1: Basic topology of a multi-channel relaying system](image-url)
Chapter 2. Background and Related Work

Source-Destination Pair

The source and destination in a practical multi-channel relaying scenario can be regarded as the OFDM transmitter and receiver in a point-to-point system, respectively. Figure 2.2 depicts a typical structure of a OFDM transmitter-receiver pair.

The serial data to be transmitted is first converted to a block of parallel data \( s = [s_1, \cdots, s_N]^T \) of dimension \( N \), over \( N \) channels. Each data stream \( s_k \) is individually modulated resulting the complex vector \( x = [x_1, \cdots, x_N]^T \). Note that various modulation levels and forms (QPSK and PAM) are possible in principle on each channel and according to the channel quality, the best one is selected for modulation. Next, the vector of data symbols \( x \) is first boosted with the power coefficient vector \( d^s = [d^s_1, d^s_2, \cdots, d^s_N]^T \), where \( d^s_k = \sqrt{P_{sk}} \) with \( P_{sk} \) being the power allocated to channel \( k \). Then the amplified data symbols goes through an inverse FFT (IFFT) block for orthogonal waveform modulation. The output of IFFT, \( X \), is mathematically given by

\[
X = [X_1, \cdots, X_N]^T = W \underbrace{\text{diag}[d^s_1, d^s_2, \cdots, d^s_N]}_{D^s} x,
\]

where \( \text{diag}[\cdot] \) is the diagonal matrix operator, and \( W \) is an \( N \times N \) IFFT matrix with sampling rate \( N/T = 1/T_s \). Here \( T \) and \( T_s \) are the OFDM symbol period and the single subcarrier period, respectively.

After the IFFT operation, an important step is to append a cyclic prefix, which is the last \( L \) samples of IFFT output, to the beginning of the IFFT output data vector \( X \). This is to avoid the inter-symbol interference (ISI) caused by multipath propagation. For this purpose, the length of cyclic prefix is chosen to be longer than the longest channel delay spread. The parallel output is then converted back to the serial, yielding an OFDM symbol with time domain representation \( [X_{N-L+1}, \cdots, X_N, X_1, \cdots, X_N]^T \). Following by pulse shaping, the resultant signal is transmitted to both the destination and the relay.

The multi-channel receiver at the destination reverses the process using FFT operation. In
particular, given the time and frequency synchronization, first the cyclic prefix is removed after serial-to-parallel block. This results in an ISI-free output $Y = [Y_1, \cdots, Y_N]^T$ given by [3]

$$Y = WH\text{diag}[H(1), \cdots, H(N)]D_s x,$$

(2.2)

where $H(1), \cdots, H(N)$ are the DFT of the channel responses, $H(i) = \sum_{l=0}^{L-1} h_l e^{-j2\pi(i-1)l/N}$. $Y$ is next fed to the FFT block for orthogonal demodulation, i.e.,

$$y = W^H Y = HD_s x.$$  

(2.3)

The output of FFT of $N$ channels are demodulated and further processed by the destination.

It is worth noting that the multi-channel destination may receive the signal both from the relay and the source over two phases. In that case, it applies the maximum ratio combining (MRC) technique to coherently add the signals to boost the overall signal-to-noise (SNR). More details are provided in subsection 2.1.4.

![Source-destination pair in multi-channel (OFDM) system](image)

Figure 2.2: Source-destination pair in multi-channel (OFDM) system

**Relay Node**

A multi-channel relay comprises a multi-channel receiver and a multi-channel transmitter with two additional blocks in between as shown in Fig. 2.3. These blocks that, together with power allocation strategy, are of our interest are the relaying operation and channel pairing blocks. We will illustrate them in the ensuing sections along with the power allocation strategies.
A relay can operate in either half-duplex or full-duplex mode. In full-duplex mode, the relay can receive and transmit at the same time on the same frequency as opposed to the half-duplex mode that allows only one at a time. To realize full-duplex relaying, recent research from both academia and industry R&Ds have made progress on building such system (also in antenna technology for Tx-Rx isolation). This is reflected in the LTE-A system development, where full-duplex relaying, referred as type 1b relaying mode, has been included in the standard [24] (page 684). For antenna isolation, some techniques such as self-signal cancellation, adaptive beam-forming, or interference rejection combining [24] that takes into account the direction of arrival signal can be used.

![Image of OFDM relay illustration](image-url)

**Figure 2.3: OFDM relay illustration**

### 2.1.2 Relaying Operation and Strategies

In this thesis, we mainly consider two types of relaying strategies: amplify-and-forward (AF) and decode-and-forward (DF), although our analysis in some cases can be extended to the other strategies as well. An AF relay amplifies the signal received from an incoming channel and directly forwards it to the next block and eventually to the destination over an outgoing channel. On the other hand, a DF relay first attempts to decode the received signal from the source over each incoming channel and then forwards a version of the decoded data on an outgoing channel to the destination.

Between DF and AF relay, one may perform better than the other depending on the situ-
ation and the system setup. Although DF has received more attention in the standardization community and technical reports for implementation, there exist several work as well that have experimentally implemented AF relaying [25] [26] (and the references therein). For instance, in [26] a test-bed of an AF OFDM-based network was constructed using distributed Alamouti transmit diversity scheme. The authors analytically and experimentally show that AF protocol not only achieves lower error rate compared to other non-relaying schemes but it is also robust against carrier frequency offsets at the relay.

Both AF and DF relays provide various performance enhancements (e.g. coverage and capacity) so that they are included in the 4th generation OFDMA system implementation [27] and [24] (chapter 30).

Once the full knowledge of channel is available at the relay, DF relaying operation seems more appealing. However, in other scenarios where not enough transceiver processing including channel estimation, encoding/decoding is accessible, AF relaying is more attractive and perhaps is the only option.

2.1.3 Channel Pairing (CP) and Power Allocation (PA)

Channel Pairing: It is conducted by the relay in which each incoming channel is matched with an outgoing channel in the channel pairing block. As different channels exhibit various qualities, a judicious CP scheme can potentially lead to significant improvement in system spectral efficiency. We say path \((m, n)\) is selected if incoming channel \(m\) is paired with outgoing channel \(n\). We further define a binary pairing indicator

\[
\phi_{mn} \in \{0, 1\}, \quad \forall m, n ,
\]

such that \(\phi_{mn}\) is 1 if path \((m, n)\) is selected and 0 otherwise. The channel pairing matrix is then defined as permutation matrix \(\Phi = [\phi_{mn}]_{N \times N}\). We consider one-to-one mappings between the incoming channels and outgoing channels. This imposes the following constraints on \(\Phi\): 

\[
\sum_{n=1}^{N} \phi_{mn} = 1, \quad \forall m; \quad \sum_{m=1}^{N} \phi_{mn} = 1, \quad \forall n .
\]
Power Allocation: It refers to determination of power coefficient $d^s$ and $d^r$ at the source and the relay respectively so as to optimize the performance metric of the system. For AF relaying, we assume that the relay first normalizes the received signal so that the average relay-transmit-power at the $n$th outgoing channel equals $P_{rn}$. Hence, if incoming channel $m$ is paired with the outgoing channel $n$, $d^r_n = \sqrt{P_{rn}}$. For DF relaying, since the received signal with unit power is first extracted at the relay, we have $d^r_n = \sqrt{P_{rn}}$.

In this thesis, we deal with individual power constraints,

\[
\sum_{n=1}^{N} P_{sn} \leq P_s, \quad \sum_{n=1}^{N} P_{rn} \leq P_r
\]  

(2.6)

and/or the total power constraint,

\[
\sum_{n=1}^{N} P_{rn} + P_{sn} \leq P_t,
\]  

(2.7)

where $P_s$, $P_r$, and $P_t$ are the maximum allowed transmission power by the source, the relay, and the combined source and relay, respectively. This is the general representation of the power limitations imposed on the system including, e.g., hardware constraint, legal or regulatory requirement, or energy conservation. Note that this general representation can be easily tailored to also specify systems with only individual power constraints, or only total power constraint, by setting one or more of $P_s$, $P_r$, and $P_t$ to sufficiently large values.

2.1.4 Input-output Relation and System Achievable Rate

We denote the channel gain over channel $k$ from source to relay, from relay to destination, and from source to destination by $h_{1k}$, $h_{2k}$, and $h_{0k}$, respectively. The received signals at the relay and destination in the first phase are given by

\[
y_r = H_1 D_s x + n_r, \quad y_d^{(1)} = H_0 D_s x + n_d^{(1)}
\]  

(2.8)

(2.9)

where $y_r = [y_{r1}, \cdots, y_{rN}]^T$ and $y_d^{(1)} = [y_{d1}^{(1)}, \cdots, y_{dN}^{(1)}]^T$ are the received signal vector at relay and destination, respectively, $H_1 = \text{diag}(h_{11}, \cdots, h_{1N})$ and $H_0 = \text{diag}(h_{01}, \cdots, h_{0N})$ are the
corresponding channel matrices, and $D_s = \text{diag}(d^s)$ with $d^s = [d^s_1, \ldots, d^s_N]^T$ being the power coefficient vector as previously indicated. The signals are assumed to be i.i.d. with average unit power $E[xx^H] = I_N$. Moreover, $n_r = [n_{r1}, \cdots, n_{rN}]^T$ and $n_d^{(1)} = [n_{d1}^{(1)}, \cdots, n_{dN}^{(1)}]^T$ are additive white Gaussian noise (AWGN) at the relay and the destination, with $n_r \sim \mathcal{CN}(0, \sigma^2_r I_N)$ and $n_d^{(1)} \sim \mathcal{CN}(0, \sigma^2_d I_N)$, respectively.

In the second phase and in the case of AF relaying, the signal received at relay $y_r$ after normalization is directly fed to the CP permutation matrix $\Phi$. The received signal vector at the destination is given by

$$y_d^{(2)AF} = H_2 D_r \Phi y_r + n_d^{(2)}$$

where $y_d^{(2)AF} = [y_{d1}^{(2)}, \cdots, y_{dN}^{(2)}]^T$, $H_2 = \text{diag}(h_{21}, \cdots, h_{2N})^T$, and $n_d^{(2)} = [n_{d1}^{(2)}, \cdots, n_{dN}^{(2)}]^T \sim \mathcal{CN}(0, \sigma^2_d I_N)$. The power coefficient vector for the processed signal at the relay is denoted as $d^r$, and we have $D_r = \text{diag}(d^r)$.

On the other hand, for DF relaying, the received signal $y_r$ is decoded to extract the original signal $x$ prior to feeding into the CP permutation matrix. Given this case, the received signal at the destination is obtained as

$$y_d^{(2)DF} = H_2 D_r \Phi x + n_d^{(2)}$$

(2.11)

In a relay system, the direct path between the source and destination may or may not be available. When the direct path is available, we assume the receiver uses maximum ratio combining to improve the reception and maximize the received SNR. This appears in the rate function.

We first consider the achievable rate obtained in the AF relay OFDM system. We can rewrite the end-to-end system equation in the following general form

$$y^{AF} = \begin{bmatrix} y_d^{(1)} \\ y_d^{(2)AF} \\ y_d^{(2)DF} \end{bmatrix} = H_{eq}^{AF} x + n_{eq},$$

(2.12)
where $H_{eq}^{AF}$ and $n_{eq}$ are the equivalent channel matrix and the equivalent noise term, respectively. For the AF relaying they are calculated as

$$
H_{eq}^{AF} = \begin{bmatrix}
H_0 \\
H_2 D_r \Phi H_1 D_s
\end{bmatrix},
$$

$$
n_{eq} = \begin{bmatrix}
n_d^{(1)} \\
H_2 D_r \Phi n_r + n_d^{(2)}
\end{bmatrix}.
$$

(2.13)

It is clear that $H_{eq}^{AF}$ as well as $n_{eq}$ are functions of the CP matrix $\Phi$, the source and relay power coefficients $D_s$ and $D_r$. The achievable rate for the AF relay multi-channel described above is given by

$$
C^{AF} = \frac{1}{2} \log \det(I + R_n^{-1}H_{eq}^{AF}(H_{eq}^{AF})^H)
$$

(2.14)

where $R_n = \mathbb{E}[n_{eq}n_{eq}^H]$ is the covariance matrix of the equivalent noise term. The factor $1/2$ reflects the half-duplex operation. Recalling $d_{sm} = \sqrt{P_{sm}}$ for all $m = 1, \ldots, N$, and $d_{rn}^n = \sqrt{\frac{P_{rn}}{P_{sm}|h_1 m|^2 + \sigma^2_r}}$ provided that $\phi_{mn} = 1$, the rate formula (2.14) can be rewritten in a non-matrix form as

$$
C^{AF} = \frac{1}{2} \sum_{m=1}^{N} \sum_{n=1}^{N} \phi_{mn} \log \left(1 + \frac{P_{sm}a_m b_n}{1 + \frac{P_{sm} a_m}{P_{rn} b_n} + P_{sm} c_m}\right),
$$

(2.15)

where $a_m = \frac{|h_1 m|^2}{\sigma^2_r}$, $b_n = \frac{|h_2 n|^2}{\sigma^2_d}$, and $c_m = \frac{|h_0 m|^2}{\sigma^2_d}$ are normalized channel power gains against the noise variance at the relay and the destination.

Next, we discuss the rate achievable in a DF relay OFDM system. The DF relaying rate depends on the decoding scheme in use at the relay. Consider the conventional repetition-coding based DF relaying [28][4], where the relay is required to fully decode the incoming message, re-encode it with repetition coding, and forward it to the intended user. The maximum achievable source-destination rate is given by [4]

$$
C^{DF}(\Phi, D_s, D_r) = \frac{1}{2} \sum_{m=1}^{N} \sum_{n=1}^{N} \phi_{mn} \min \{\log(1 + a_m P_{sm}), \log(1 + c_m P_{sm} + b_n P_{rn})\}.
$$

(2.16)
2.1.5 Channel Pairing vs. Bit Piping in DF Relay

For DF relaying one may adopt the bit piping scheme instead of channel pairing. Given the total power budget of each hop, bit piping maximizes the rate of each hop separately over their respective channels’ power. Therefore, with no direct link (DL) available, the minimum of the first hop’s rate and the second hop’s rate yields the maximum achievable rate in this scheme. With DL available, the total achievable rate depends on the coding scheme being used, and indeed not every coding technique at the relay can utilize DL in bit piping scheme.

Throughout this thesis, we make use of repetition code [28][4] as a practical and basic coding technique for the DF relay. Under this coding technique, bit piping has to ignore the DL since the destination is unable to specify the channels of the direct and relay links that carry the same message to be able to apply the MRC technique. On the other hand, in pairing it is clear which channels carry which messages, thus the pairing scheme can exploit the DL along with the relay link using MRC technique. We implicitly consider the single-block decoding scheme at the destination that is robust against error propagation. The multi-block decoding schemes such as backward decoding [29][30] or window decoding [31][32] despite achieving higher rate by utilizing DL are subject to error propagation. Moreover, the backward decoding scheme also incurs substantial decoding delay.

In what follows we conduct a simple numerical simulation to show that indeed pairing outperforms piping in some cases for DF relaying when the repetition coding is adopted. We consider a 8-channel dual-hop system with a source (S), a half-duplex relay (R) and a destination (D) forming an isosceles triangle with the DF relay in the middle. Fig. 2.4 depicts the end-to-end normalized achievable rate vs. the relay-node angle whose variation indicates the S-D distance variation when keeping the S-R and R-D distances fixed. The nominal SNR defined as the ratio between the source power and the S-D path-loss plus the noise variance, is set to 6dB, and the relay total power is assumed to be the same as the source total power. As can be observed, the joint pairing and PA technique yields almost 10% higher rate than the piping scheme does. We have also plotted the rate obtained from direct communication where
no relay is involved and communication occurs in one-time slot. As expected, when source is sufficiently close to the destination, \( \theta < \frac{2\pi}{3} \), direct communication achieves a higher rate comparing with piping and pairing schemes. However, when they are far from each other, pairing jointly with power allocation achieves the highest rate.

In general, relaying is superior over the direct communication once the additional power contributed by the relay overcomes the performance loss due to the two-time slot communication in half-duplex relaying. For the full-duplex relaying, however, this loss does not exist, and therefore the superiority of relaying strategy always holds.

Noteworthy is that for other scenarios such as AF relaying or multi-source multi-destination cases bit piping is not doable while pairing is always possible.

So far, we have illustrated the basic structure of a multi-channel system and discussed its general input-output relations and the achievable rates for both AF and DF case. In the upcoming chapters, we consider various practical and complicated scenarios which essentially build upon the basic multi-channel relaying system introduced in Section 2.1.1. For these multi-channel relaying scenarios, our main objective is to find the optimal channel pairing.
and power allocation matrices that maximizes their respective system metrics. In doing so, we extensively make use of optimization techniques including both convex and non-convex methods. Therefore, we devote some space in this chapter to introducing some techniques that we often employ to find the optimal system solutions.

## 2.2 Optimization Tools and Techniques

Optimization problems arise in a variety of disciplines such as automatic control systems, estimation and signal processing, communications and networks, etc. In general, they can be classified as the convex and non-convex optimization problems. For convex problems, there exists several efficient algorithms that produce the globally optimal solution. In fact, with the recent improvements in computing and optimization theory, convex optimization can be solved almost as easy as a linear programming problem. As a result, it is quite desirable to deal with a convex problem in comparison with a non-convex problem whose solution is usually NP-hard. For this reason, we always attempt to convert a non-convex problem to a convex problem knowing that this conversion may not be always possible.

### 2.2.1 Convex Optimization

As stated, recognizing or formulating a problem as a convex optimization problem has many benefits. The most notable one is that the problem can then be solved, very reliably and efficiently, using interior-point methods or the other algorithms designed specifically for convex problems.

An optimization problem in general is expressed as

\[
\begin{align*}
\min_{\mathbf{x}} & \quad f_0(\mathbf{x}) \\
\text{s.t} & \quad f_i(\mathbf{x}) \leq 0, \quad i = 1, \cdots, k, \\
& \quad h_j(\mathbf{x}) = 0, \quad j = 1, \cdots, l,
\end{align*}
\]  

(2.17)
where \( x \in \mathbb{R}^n \) is the optimization variable vector. We denote by \( p^* \) the optimal value of (2.17), which is the so called primal value. The optimization problem in (2.17) is convex if

- The functions \( f_0, \cdots, f_k : \mathbb{R}^n \rightarrow \mathbb{R}^1 \) are convex satisfying the convexity condition

\[
    f_i(\alpha x + (1-\alpha)y) \leq \alpha f_i(x) + (1-\alpha)f_i(y) \quad i = 1, \cdots, N
\]

(2.18) for all \( \alpha \in \mathbb{R} \) and all \( x, y \in \mathbb{R}^n \).

- The equality constraint functions \( h_j(x) \)s must be affine, i.e., \( h(x) = a_j^T x - b_j \), where \( a_j \) and \( b_j \) \( j = 1, \cdots, l \) are a set of vectors and scalar constant, respectively. This is a stronger condition than just being convex.

If \( f_i \) is twice differentiable, the convexity condition implies the following condition: \( f_i \) is convex if and only if domain of \( f_i \) is a convex set and its Hessian matrix is positive semidefinite, (Section 3.1.4 of [33]) denoted as

\[
    \Delta^2 f_i(x) \succeq 0.
\]

(2.19)

Matrix \( A \) is positive semidefinite if and only if for any vector \( x \), \( x^T A x \geq 0 \), or equivalently all the eigenvalues of \( A \) are non-negative.

If minimization is replaced by maximization in (2.17), for convexity the negative function \( -f_0 \) needs to be convex, or equivalently \( f_0 \) must be concave. A concave function \( f_0 \) must satisfy (2.18) with the backward inequality, and if it is twice differentiable, its Hessian matrix must be negative semidefinite.

For better illustration, we consider the AF rate function in (2.15) of path \((m, n)\) as an example. By a quick inspection on its Hessian matrix, one can realize that this function is neither convex nor concave in \((P_{sm}, P_{rn})\) for given \( a_m, b_n \) and \( c_m \). However, the upperbound approximation of this rate, i.e.,

\[
    C_{mn}^{AF} \approx \frac{1}{2} \log \left( 1 + \frac{P_{sm}a_m P_{rn}b_n}{P_{sm}a_m + P_{rn}b_n + P_{sm}c_m} \right)
\]

(2.20)

is easily proven to be concave.
There are some operations that preserve the convexity (concavity) of functions. Simple operations such as addition, scaling or pointwise maximum (minimum) are some examples of such operations. In particular, if \( f_1 \) and \( f_2 \) are concave functions, their minimum function \( f_{\min}(x) = \min(f_1(x), f_2(x)) \) is also a concave function. This follows that the DF rate function in (2.16) is concave since each term inside the min function is concave.

We note that any maximization problem can be easily transformed into a minimization problem by considering the negative of the objective function. Hence, the following discussion with slight change can be applied to maximization problems.

**Duality**

Duality is important because apart from giving fundamental insights about the problem, it either leads to the optimal solution for convex problems, or results in a lower bound for non-convex problems. We often convert the primal (original) optimization problem to its dual, mainly in the form of a *Lagrangian* duality.

*Lagrange and Dual Functions:* The Lagrangian duality incorporates the constraints in (2.17) by augmenting the objective function with a weighted sum of the constraint functions. The Lagrange function of (2.17) is defined as

\[
L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{k} \lambda_i f_i(x) + \sum_{j=1}^{l} \mu_j h_j(x),
\]  

(2.21)

where \( \lambda_i \) and \( \mu_j \) are referred as the Lagrange multipliers corresponding to the \( i \)th inequality constraint \( f_i(x) \), and \( j \)th equality constraints, \( h_j(x) \), respectively. The dual function is defined as

\[
g(\lambda, \mu) = \min_x L(x, \lambda, \mu),
\]  

(2.22)

which is convex in \((\lambda, \mu)\) since it is the pointwise minimum of a set of affine functions of \((\lambda, \mu)\). It is not difficult to see that for each pair \((\lambda, \mu)\) with \( \lambda \succeq 0 \), the dual function \( g(\lambda, \mu) \) gives us a lower bound on \( p^* \), the optimal value of the original optimization problem (2.17).
Dual Problem, Weak and Strong Dualities: The dual problem associated with the optimization problem (2.17) is given by

$$\max_{\lambda, \mu} \ g(\lambda, \mu)$$

$$\text{s.t} \quad \lambda \succeq 0,$$  \hspace{1cm} (2.23)

where $\succeq$ is an element-wise inequality. Let $d^*$ denote the optimal value of (2.23). This value can be efficiently calculated since the dual problem (2.23) is convex. Moreover, we note that $d^*$ is the best lower bound of $p^*$ obtained from the dual function. This fact leads to a weak duality property which states: For any optimization problem $d^* \leq p^*$ holds. The difference $p^* - d^*$ is referred as the duality gap.

Strong duality calls for a more stricter condition $d^* = p^*$ or zero duality gap. Obviously, this condition fails to prevail for all optimization problems, however, it is shown that we have the strong duality for the convex problem [33]. Convexity is a sufficient condition for strong duality, and a more general sufficient condition is called time sharing as it featured in [34]. We will provide more details on time sharing in the next chapters.

When strong duality holds, the primal optimal solution $x^*$ and dual optimal $(\lambda^*, \mu^*)$ must satisfy the following conditions which are known as KKT conditions

$$\lambda^*_i f_i(x^*) = 0 \quad i = 1, \ldots, k, \hspace{1cm} (2.24)$$
$$h_j(x^*) = 0 \quad j = 1, \ldots, l \hspace{1cm} (2.25)$$
$$\lambda^* \succeq 0, \hspace{1cm} (2.26)$$

and if $\{f_i(x)\}$ and $\{h_j(x)\}$s are differentiable,

$$\Delta f_0(x^*) + \sum_{i=1}^{k} \lambda^*_i \Delta f_i(x^*) + \sum_{j=1}^{l} \mu^*_j \Delta h_j(x^*) = 0. \hspace{1cm} (2.27)$$

Unfortunately, most of the optimization problems under investigation in this thesis are not convex in their original form. Due to the combinatorial nature of the channel pairing, they are often in integer programming form which is an important class of non-convex problems.
2.2.2 (Mixed) Integer programming

Integer programming (IP) deals with the set of optimization problems in which all of the variables are restricted to be integer. Almost similarly, mixed integer programming requires only some of the variables and not all to be integer. Integer programming is one of the exciting and fast developing areas in optimization theory and operation research. To date, there is no general polynomial-time solutions known for this set of problems. For some IP problems with special structure, however, the polynomial time solution exists.

Assignment problems (AP) are a well known example of Integer programming problems whose solution can be found efficiently in a polynomial time. As we will make use of it in the later chapter, we briefly illustrate it below.

Assignment Problem

An assignment problem optimally matches the elements of two or more sets of elements in a one-to-one manner. Examples arise in many practical problems such as matching tasks to the agents, or matching teachers to students and to projects. The dimension of an AP refers to the number of sets involved in matching. For instance, the dimension in our former example is 2 while that in our latter example is 3. A two-dimensional (2-D) AP, as the classical AP, is mathematically stated as

$$\min_y \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} y_{ij}$$

s.t

$$\sum_{i=1}^{N} y_{ij} = 1 \quad \forall j$$

$$\sum_{j=1}^{N} y_{ij} = 1 \quad \forall i$$

$$y_{ij} \in \{0, 1\} \quad \forall i, j,$$

where $a_{ij}$ is the cost of matching element $i$ from set $A$ with element $j$ from set $B$. Constraints (2.29) and (2.30) ensure the one-to-one mapping. Since the constraint matrix of this problem
is totally unimodular [20], it is shown that the continuous relaxation of this problem, which is a linear programming problem, leads to the integer solution. Hence, any linear programming approach can produce the desirable integer solution in polynomial time. There are also some efficient algorithms specific for two-dimensional AP problems such as Hungarian algorithm [35] whose complexity is in the order of $O(N^3)$.

Unlike a 2-D AP, which has a polynomial solution, three or higher dimensional APs are in general NP hard. In the case of three-dimensional (3-D) AP, two models are defined: Axial 3-D AP and planar 3-D AP. In what follows, we describe the axial 3-D AP which resembles in many aspect to the classical 2-D AP but with NP-hard solution. This problem as, integer linear programming, is written as

$$
\min \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} a_{ijk}y_{ijk} \quad (2.32)
$$

$$
s.t \sum_{i=1}^{N} \sum_{k=1}^{N} y_{ijk} = 1 \quad \forall j \quad (2.33)
$$

$$
\sum_{i=1}^{N} y_{ijk} = 1 \quad \forall i \quad (2.34)
$$

$$
\sum_{j=1}^{N} y_{ijk} = 1 \quad \forall k \quad (2.35)
$$

$$
y_{ijk} \in \{0, 1\} \quad \forall i, j, \quad (2.36)
$$

where analogous to the 2-D case, $a_{ijk}$ is the cost associated with matching elements $i$ from set $A$ with elements $j$ from set $B$ and together assigned to element $k$ from set $C$. Sets $A$, $B$ and $C$ can be regarded as sets of jobs, workers and machines, respectively. $y_{ijk}$ would be then an indicator variable for job $i$ done by worker $j$ with machine $k$. Constraints (2.33)-(2.35) restrict that each element of a group is matched with only one pair of elements of the other two sets. Two methods are generally used to tackle the 3-D AP. One is the branch and bound algorithm which literally enumerates all basic feasible points. The complexity of this algorithm is not necessarily polynomial. On the other hand, the second approach is through using duality to provide an easily computed lower bound.
Despite being NP-hard in general, some 3-D AP with special structure may have a polynomial-time solution. For instance, suppose the cost of matching job $i$ to worker $j$ together to machine $k$ be additive in a sense that it can be expressed as $i.e., a_{ijk} = b_{ij} + c_{jk}$, where $b_{ij}$ is the cost of assigning job $i$ to worker $j$ and $c_{jk}$ is the cost of assigning worker $j$ to machine $k$. Then it is shown that the corresponding 3-D AP can be converted to a 2-D classical AP with polynomial solution [36]. We will also deal with a 3-D AP in Chapter 4 for which we can present an optimal solution using the reformulation technique.

**2.2.3 Dual Decomposition and Subgradient Algorithm**

We employ the dual decomposition and subgradient algorithm quite frequently in different parts of our thesis owing to the special structure of the optimization problems arising in the system models. We present the concept of dual decomposition by using the following example. Consider the underlying optimization problem over variable $x = [x_1, x_2]$

$$\min_{x_1, x_2} f_1(x_1) + f_2(x_2) \quad (2.37)$$

$$s.t \quad g_1(x_1) \leq P_1, \quad (2.38)$$

$$g_2(x_2) \leq P_2, \quad (2.39)$$

$$h_1(x_1) + h_2(x_2) \leq P_3, \quad (2.40)$$

where (2.40) is the only coupled constraint on $x_1$ and $x_2$ together. Before proceeding and forming the Lagrange function, it is worth noting that it is perfectly allowed to dualize and include only a subset of constraints, and not necessarily all the constraints, in the Lagrange function. We would just need to carry the rest of the constraints during the process of dual maximization. Further details can be found in the Implicit and Explicit Constraints section (P.134) of [33]. One can readily provide a proof for the zero-duality gap in such dualization in case of convexity.
As a result, we define the Lagrange function as \( L(\lambda, x_1, x_2) = f_1(x_1) + f_2(x_2) + \lambda^T (h_1(x_1) + h_2(x_2) - P_3) \), and correspondingly the dual function as

\[
g(\lambda) = \min_x L(\lambda, x_1, x_2) \tag{2.41}
\]

\[
s.t \quad g_1(x_1) \leq P_1,
\]

\[
g_2(x_2) \leq P_2.
\]

It is clear that minimization of \( g(\lambda) \) in (2.41) over \( x \) can be decomposed into two independent subproblems to find the optimal \( x_1 \) and \( x_2 \), i.e.,

**Subproblem 1:**

\[
\min_{x_1} f_1(x_1) + \lambda^T h_1(x_1) \tag{2.42}
\]

\[
s.t \quad g_1(x_1) \leq P_1,
\]

**Subproblem 2:**

\[
\min_{x_2} f_2(x_2) + \lambda^T h_2(x_2) \tag{2.43}
\]

\[
s.t \quad g_2(x_2) \leq P_2.
\]

Since \( P_3 \) is constant, it can be removed from the optimization. This dual decomposition often leads to a significant search space reduction, enabling software to find the dual solution much faster. It is evident that the decomposition above can be extended to other problems with general objective function \( \sum_i f_i(x_i) \) and coupled constraint \( \sum_i h_i(x_i) < P \). As we will see later, this method is applicable to the rate maximization problem under total power constraint. We will provide more details as we use this technique in the upcoming chapters.

Next we explain the subgradient method [37] that is widely used for maximizing (minimizing) the dual function. For illustration purposes, we again use the optimization problem in (2.40) and its corresponding dual problem as an example. Recalling the dual problem, we want
to find optimal $\lambda$ that maximizes $g(\lambda)$ subject to $\lambda \geq 0$. To maximize $g(\lambda)$, the subgradient method uses the iteration

$$\lambda^{(k+1)} = \lambda^{(k)} + \alpha_k d^{(k)} T,$$  \hspace{1cm} (2.44)

where $\lambda^k$ is the $k$th iteration of Lagrange multiplier, $d^k$ is any subgradient of $g$ at $\lambda = \lambda^k$, and $\alpha_k$ is the $k$th step size in the direction of the subgradient. Recalling the subgradient’s definition, at each iteration we need to find $d$ of $g(\lambda)$ at $\lambda$ such that $g(\lambda') \leq g(\lambda) + d^T (\lambda' - \lambda)$ for all $\lambda'$. Let $x_1^*$ and $x_2^*$ denote the optimal values obtained from optimization (2.42) and (2.43), respectively, for given $\lambda$. It is then shown that $d = h_1(x_1^*) + h_2(x_2^*) - P_3$ is the subgradient of $g(\lambda)$.

The subgradient method is mainly used for non-differentiable functions as opposed to the gradient method that is designed solely for differentiable functions. They have some notable differences in step size and the convergence behavior. For instance, the subgradient method is not a descent method in a sense that the function value does not necessarily increase over each iteration. Therefore it is common to keep track of the best point, one with the largest $g(\lambda)$. Furthermore, the subgradient uses the pre-calculated step-size for convergence. There are various types of step size under which the iteration in (2.44) is guaranteed to converge to the optimal $\lambda^*$ under some general conditions [37][38]. We specifically use the constant step size rule i.e., $\alpha_k = \alpha$, or the non-summable, square summable rule which requires the step size to be nonsummable $\sum_k \alpha_k = \infty$ but square summable $\sum_k [\alpha_k]^2 < \infty$. We will elaborate the advantages of either these rules in the next chapters where we invoke them.

2.3 Related Works

Much research effort has been devoted in the past to understanding the behaviors and benefits of cooperative relaying. Some early works include information theoretic studies following the seminal work in [28], and extensive studies have been conducted in network-layer designs focused on packet forwarding in ad hoc and mesh networks [39]. More recently, the benefit of
cooperation diversity has been identified from a different perspective, in the new paradigm of cooperation diversity at the physical layer, as a means to improve reception [4][5]. This paradigm has its root in dual-hop relaying, and it has generated much interest in further analyzing the cooperative gain in a variety of dual-hop relay channels and in how to realize such gain with practical schemes [40][41].

2.3.1 Resource Allocation for Single-channel Relaying

Many studies concern optimal power allocation (PA) between the source and the relay in a single-relay cooperation under total power constraint, where the performance metrics may be data rate [42][43], outage probability [44][45], or bit error rate [46]. Assuming full channel state information is not attainable, the authors of [47] considered a MIMO-AF relay system and proposed a beamforming codebook based on Grassmannian manifold.

In a multi-relay setup with the objective of data-rate maximization, the problem of PA for relay cooperation have been considered in [48][49][50] and for distributed relay selection schemes are studied in [51] and [52]. The authors of [53] and [54] adopted the maximization of the system lifetime as the objective metric and optimized the relays’ power. They essentially argued that for the networks with battery-operated relays, minimizing total transmission power per usage does not necessarily indicate the lifetime maximization. In [53] and [54] the lifetime was defined as the time duration for which the network is able to sustain a certain QoS requirement.

2.3.2 Resource Allocation without Channel Pairing for Multi-channel Relaying

Below, we will separately review the prior works on resource allocation without CP in multi-channel relaying with single-destination (single-user system) and with multiple-destination (multi-user system).
CHAPTER 2. BACKGROUND AND RELATED WORK

Single-user System

For a dual-hop OFDM system with single destination the problem of optimal PA was studied by many [6][7][8][9] for different relay strategies and power constraints. The authors of [8] and [7] proposed a PA solution for AF and DF relaying strategies, respectively, to maximize the total end-to-end rate. With the same objective, a multi-relay extension of the previous works was considered in [55] and a joint PA and relay selection solution was developed. In [6], the authors considered a DF multi-relay setup and proposed a number of bit and power allocation schemes aiming at minimizing the total transmission power under a predefined rate constraint. Beside PA problem, the problem of bandwidth allocation and relay strategies was considered in [56] in a multi-channel multi-relay scenario. In particular, they proposed an optimization framework to maximize the system total utility based on dual decomposition and time sharing property. For a multi-hop OFDM-based network [57] studied the problem of power and time allocation under a long-term total power constraint.

Unlike above works which all assumed the knowledge of global channel gains, in [9] the authors adopted a limited feedback system and assumed only a limited information of channels can be available at the source/relay. They proposed a PA scheme that uses a codebook of quantized PA vectors designed offline under two design criteria, maximizing capacity and minimizing error rate.

Multi-user System

The papers below have studied the problem of resource allocation in a multi-channel multi-destination (multi-user) case. In a multi-user communication environment, beside PA and possibly CP, a subset of incoming-outgoing channel pairs are assigned to a user for data transmission. This channel-user assignment provides additional degree of freedom which can be exploited to offer better diversity and overall system performance. Studies on joint PA, CP, and channel-user assignment in multi-destination scenarios have so far been scarce. Due to its combinatorial nature which entails a mixed integer programming problem, previous attempts
often focused on optimizing of subset of these three resources [10][11][12][58][13][14]. Several suboptimal approaches providing reduced complexity were also proposed under certain system models [11][14][13]. Without PA and CP, the authors of [10] considered the problem of channel-user assignment for a multi-destination multi-relay setup. They suggested a graph theoretical approach to solve the 3-D resulting integer programming problem.

**Joint Power Allocation and Channel-user Assignment**: More attention in the literature has been placed on joint PA and channel-user assignment. This joint problem was studied in [11] for a single-relay AF multi-destination scenario and a suboptimal scheme was proposed to maximize the weighted-sum rate. This study was generalized to multi-relay multiple source-destination pairs scenario in [14], [13] and [15]. Apart from PA and channel-user assignment, [14] also addressed the time scheduling problem to maximize the sum rate under a TDMA multi-cell framework with inter-cell interference. They proposed a two-stage suboptimal algorithm which alternatively optimizes channel-user assignment and time scheduling at one stage and PA at the other stage. A similar two-stage PA and channel-user assignment approach was suggested in [13] to maximize the total rate under minimum users’ rate constraint to ensure the fairness. In [15], the authors considered the problem of transmission strategy selection as to whether use the DF relay or not, and then they propose some suboptimal approach to maximize the minimum rate among the pairs.

### 2.3.3 Resource Allocation with Channel Pairing for Multi-channel Relaying

We again break the previous studies on this topic to two single-user and multiple-user cases:

#### Single-user Systems

For an OFDM system, the concept of CP was first introduced independently in [59] and [60] for a dual-hop AF relaying system where heuristic algorithms for pairing based on the order of
channel quality were proposed. For relaying without the direct source-destination link available, [59] used integer programming to find the optimal pairing that maximizes the sum SNR. From a system-design perspective, the sorted-SNR CP scheme was proposed in [60] and was shown optimal for the noise-free relaying case, under the assumption of uniform PA.

These works sparked interests for more research in this area. In the absence of the direct source-destination link, for the practical case of noisy-relay, by using the property of L-superadditivity of the rate function, the authors of [61] proved that the sorted-SNR CP still remains optimal for sum-rate maximization in dual-hop AF relaying OFDM system. Subsequently, it was further proved in [62], through a different approach, that the sorted-SNR CP scheme is optimal for both AF and DF relaying in the same setup. When the direct source-destination link is available, [63] presented two suboptimal CP schemes. For the same setup, in this proposal, we propose a low complexity optimal CP scheme for dual-hop AF relaying, and characterize the effect of direct path on the optimal pairing. In addition, we show that under certain conditions on relay power amplification, among all possible linear processing at the relay, the CP is optimal. Nonetheless, jointly optimizing of PA and CP have not been addressed in any of the works above.

**Joint Power Allocation and Channel Pairing**: The problem of jointly optimizing CP and PA was studied in a dual-hop OFDM system for AF and DF relaying in [64] and [65], respectively, where the direct source-destination link was assumed available. The joint optimization problems were formulated as mixed integer programs and solved in the Lagrangian dual domain. Exact optimality under arbitrary number of channels was not established. Instead, by adopting the time-sharing argument [34] in their systems, the proposed solutions were shown to be optimal in the limiting case as the number of channels approaches infinity.

Without the direct source-destination link, jointly optimizing CP and PA for DF relaying in a dual-hop OFDM system was investigated in [66] and [67], where [66] assumed a total power constraint shared between the source and the relay, and [67] considered individual power constraints separately imposed on the source and the relay. In both cases, two-step separate CP
and PA schemes were proposed and then proved to achieve the jointly optimal solution. For this dual-hop setup, it was shown that the optimal CP scheme is the one that maps the channels solely based on their channel gains independent of the optimal PA solution.

The authors of [68] proposed an adaptive PA algorithm to maximize the end-to-end rate under the total power constraint in a multi-hop OFDM relaying system. For a similar network with DF relaying, [57] studied the problem of joint power and time allocation under the long-term total power constraint to maximize the end-to-end rate. Furthermore, in [68], the idea of using CP to further enhance the performance was mentioned in addition to PA. However, no claim was provided on the optimality of the pairing scheme under the influence of PA. The optimal joint CP and PA solution remained unknown. In the next chapter of this thesis, we basically address this problem.

Multi-user Systems

**Joint Power Allocation and Channel Pairing**: In [12] and [58], the authors studied the joint optimization of PA and CP for a multi-channel multi-destination case under the predetermined channel-user assignment. While the former adopted DF strategy and aimed at maximizing weighted-sum rate as the goal, the former focused on AF relaying aiming at minimizing total power. In [58], the authors additionally provided a heuristic algorithm to address the general question of "who helps whom".

**Joint Power Allocation, Channel Pairing, and Channel-user Assignment**: To the best of our knowledge, [69] is the only study on the joint resource allocation for a multi-channel multi-destination OFDMA system. It considered DF relaying without the direct source-destination link. Under the total power constraint, a three-stage was proposed: first the relay assigns each outgoing channel to the user with the maximum SNR on that channel, then the sorted CP scheme in [62] is used, and finally the water-filling PA is applied. It was shown that this separate optimization algorithm maximizes the sum-rate in the considered system. In this thesis, we consider more general relaying strategies that uses the direct source-destination link, where
the simple sorted CP scheme is no longer optimal. Furthermore, our proposed solution framework accommodates relaying strategies other than DF, individual power constraints in addition to total power constraints, as well as weighted sum-rate and max-min fairness objectives.

2.3.4 Bit Loading for Multi-channel Point-to-Point

The resource allocation schemes proposed by all above works lead to a non-integer rate adaptation. In practice, since there is only a finite modulation level, the continuous rate adaptation is infeasible. For conventional point-to-point OFDM systems, various power allocation and bit loading schemes were proposed incorporating the discrete nature of modulation levels, with the performance metric being either total bit rate [70]-[75], or symbol error rate [76]-[77]. Other additional constraints might be involved in this resource optimization as well, such as maximum size of QAM or PAM constellations [75] or maximum allowed transmission power over each channel [78]. In existing literature, [70] was the first to consider this problem and developed a greedy algorithm that is optimal in maximizing throughput. This algorithm successively allocates bits to a channel whose incremental energy to transmit one additional bit is the least among all channels. Subsequently, optimal algorithms with reduced complexity were presented in [73] for QAM and in [75] and [74] for general modulation schemes. To further alleviate the computational burden, some less complex but suboptimal solutions were proposed in [71], [72] and [77]. The authors of [71] and [72] took advantage of the water-filling solution obtained from continuous rate maximization and then round it off to the integer rates. On the other hand, [77] adopted uniform power allocation to reduce complexity. It begins to load each channels with 64QAM and iteratively reduces the modulation size for the channels until the target mean error rate is achieved. Recognizing the necessary and sufficient condition for the optimal solution, [73] devised an optimal bit allocation algorithm with complexity $O(N)$ (as compared to $O(N^2)$ in [70]), with $N$ being the number of channels, to minimize the total transmit power subject to a target rate and a specific error probability over each channel. However, the method is only optimal for QAM modulations. In such system, the Gap approximation [79] can be
employed to formulate the rate for given error probability. For the general modulation schemes other than QAM, [75] and [74] developed optimal algorithms with complexity $O(N \log N)$. Despite being optimal and achieving the same complexity, the latter work is more favorable due to its less implementation effort. [74] essentially employs the Lagrange technique, in particular Lagrange-multiplier bisection search along with efficient look-up table, to provide a computationally-efficient solution that maximizes the integer total bit rate.

### 2.3.5 Bit Loading for Multi-channel Relaying

Studies on bit allocation for multi-channel relaying has, so far, been scarce. In [80], the authors investigated the bit loading and power allocation problem for both AF and DF relaying OFDM systems with an aim to minimize the total transmit power subject to a target bit rate and symbol error rate. Under total power constraint, they converted the problem to that of a equivalent point-to-point system, and develop the optimal solution based on the greedy approach in [70]. For a scenario with individual power constraints, a suboptimal approach was proposed assuming fixed source-relay power ratio for each channel.

For a multi-hop DF relaying system, [81] proposed a suboptimal approach for power optimization through bit loading, bandwidth and channel assignment with a minimum required rate for each hop. The greedy algorithm [70] is applied over each individual hop to meet the minimum target bit rate with minimal power.

Neither of these works have considered channel pairing. The common technique in these works is to convert the dual-hop system to an equivalent point-to-point system, where the conventional optimal greedy algorithm can be applied. However, when channel pairing needs to be jointly optimized, such technique is no longer applicable, because the conversion of the dual-hop system to an equivalent point-to-point system changes whenever a different channel pairing is used.

To the best of our knowledge, [82] is the only work considering the joint resource allocation problem consisting of channel pairing and bit and power allocation in a multi-channel system.
It focused on DF relaying and total power constraint. Aiming at minimizing the total transmit power under a target bit rate constraint, the authors proposed a suboptimal approach based on separate optimization, where pairing is conducted first and then, similar to [80] and [81], the greedy algorithm [70] is applied over the paired channels that can be viewed as an equivalent point-to-point transmission. In contrast, we present a new approach to jointly optimize channel pairing, power allocation, and bit loading. In Section 5.5, we demonstrate that our jointly optimal solution can significantly outperform many simpler suboptimal solutions, including one based on the above separate optimization.
Chapter 3

Jointly Optimal CP and PA for Multi-channel Multi-hop Relaying

This chapter investigates the problem of channel pairing (CP) and power allocation (PA) in a multi-channel multi-hop relay network to enhance the end-to-end data rate. Both AF and DF relaying strategies are considered. Given fixed power allocation to the channels, we show that channel pairing over multiple hops can be decomposed into independent pairing problems at each relay, and a sorted-SNR channel pairing strategy is sum-rate optimal, where each relay pairs its incoming and outgoing channels by their SNR order. For the joint optimization of channel pairing and power allocation under both total and individual power constraints, we show that the problem can be decoupled into two subproblems solved separately. This separation principle is established by observing the equivalence between sorting SNRs and sorting channel gains in the jointly optimal solution. It significantly reduces the computational complexity in finding the jointly optimal solution. It follows that the channel pairing problem in joint optimization can be again decomposed into independent pairing problems at each relay based on sorted channel gains. The solution for optimizing power allocation for DF relaying is also provided, as well as an asymptotically optimal solution for AF relaying.

One may view a CP scheme at each relay as a routing scheme embedded in the network.
CHAPTER 3. JOINTLY OPTIMAL CP AND PA FOR MULTI-CHANNEL MULTI-HOP RELAYING

router. However, despite bearing some resemblance, the CP problem differs from the conventional multi-channel routing problem: For channel pairing, the total cost of two paired incoming and outgoing links is not additive as it is typically assumed in the routing case. Furthermore, the cost of each link cannot be independently defined in CP. The source-destination achievable data rate is dictated by the end-to-end signal-to-noise ratio (SNR), which is a non-linear function of the channel gain and power used on each link.

The rest of this chapter is organized as follows. In Section 3.1, we present the system model and joint optimization formulation. In Section 3.2, given a fixed PA solution, we provide the optimal CP scheme based on the sorted SNR for both AF and DF strategies. The joint optimization problem of CP and PA is considered in Section 3.3, where the separation principle between CP and PA optimization is established. The optimal PA solution is then discussed in Section 3.3.3 for multi-hop relaying under both total and individual power constraints. In Section 3.3.4, we further extend the joint optimization and separation results to the multi-destination scenario. The numerical study are provided in Section 3.4.

3.1 System Model and Problem Statement

We mainly focus on an $M$-hop multi-channel relay network where a source node communicates with a destination node via $(M-1)$ intermediate relay nodes over $N$ equal-bandwidth channels, as illustrated in Fig. 3.1. Extension to the multi-destination scenario turns out to be direct and is presented in Section 3.3.4. We denote by $h_{m,n}$, for $m = 1, \cdots, M$ and $n = 1, \cdots, N$, the channel response on channel $n$ over hop $m$. The additive noise at hop $m$ is modeled as an i.i.d. zero mean Gaussian random variable with variance $\sigma_m^2$. We define $a_{m,n} \doteq \frac{|h_{m,n}|^2}{\sigma_m^2}$ as the normalized channel gain against the noise power over channel $n$ of hop $m$. In the rest of the presentation, we simply refer to it as channel gain without causing confusion. We make the common assumption that the full knowledge of global channel gains is available at a central
controller, which determines the optimal CP and PA\(^1\). We further assume that the destination is out of the transmission zone of the source, and therefore, there is no direct transmission link and the destination only receives the signal from the \((M-1)\)th relay. Similarly, it is assumed that each relay receives the signal only from the preceding relay and not from multiple relays. For \(M\)-hop relaying, a transmission from source to destination occupies \(M\) equal time slots, one for each hop. In the \(m\)th slot, \(m = 1, \ldots, M\), the \(m\)th node (the source node if \(m = 1\), otherwise the \((m-1)\)th relay node) transmits a data block to the \((m+1)\)th node (the destination node if \(m = M\), otherwise the \(m\)th relay node) on each channel. The transmission of different data blocks in different hops may occur concurrently, depending on the scheduling pattern for spatial reuse of spectrum.

![Illustration of multi-channel multi-hop relaying network with channel pairing.](image)

In practice, multi-hop relaying can be used to extend the coverage in the area remote and outside of the main and urban cities. Moreover, multi-hop relaying is employed in wireless mesh networks. Table 3.1, published by DoCoMO Co., illustrates the scenarios for which multi-hop relaying can be employed in cellular communication. Multi-hop communication becomes more important for the communication at high carrier-frequency since the signal attenuation is severe and detrimental at such frequencies. Multi-hop relaying is one way to avoid the amplitude reduction.

We consider both AF and DF relaying strategies as defined in the previous chapter for this

\(^1\)However, we show later that, for joint CP and PA optimization, the CP solution requires only local channel information at each relay, and given the proposed CP solution, a uniform PA scheme without using channel information is near optimal even at moderately high SNR for AF relaying. This is because the water-filling solution leads to the uniform PA solution at high SNR, the fact that can be seen in a point-to-point multi-channel case as well.
## Table 3.1: Relay-technology deployment scenarios [27]

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Deployment</th>
<th>Number of hops</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural area</td>
<td>Extend coverage to mountainous regions, sparsely populated areas</td>
<td>1 hop</td>
</tr>
<tr>
<td>Wireless backhaul</td>
<td>Extend coverage to mountainous regions, sparsely populated areas, remote islands</td>
<td>1 hop, multiple hops</td>
</tr>
<tr>
<td>Emergency or temporary coverage</td>
<td>Provide temporary coverage at times of disasters, events, etc.</td>
<td>1 hop, multiple hops</td>
</tr>
<tr>
<td>Urban hot spot</td>
<td>Expand coverage and enhance throughput in urban areas with high concentrations of traffic</td>
<td>1 hop</td>
</tr>
<tr>
<td>Dead spot</td>
<td>Fill coverage hole</td>
<td>1 hop, multiple hops</td>
</tr>
<tr>
<td>Indoor hot spot</td>
<td>Expand coverage to indoor environments and enhance throughput</td>
<td>1 hop</td>
</tr>
<tr>
<td>Group mobility</td>
<td>Install relay stations in public vehicles to reduce handover and location-registration control signals</td>
<td>1 hop</td>
</tr>
</tbody>
</table>
setup. The only difference is that an AF or DF relay forwards its signal to the next intermediate relay node instead of destination in the dual-hop case.

The relay also conducts CP, matching each incoming channel with an outgoing channel. We denote path $\mathcal{P}_i = (c(1,i), \cdots, c(M,i))$, where $c(m,i)$ specifies the index of the channel at hop $m$ that belongs to path $\mathcal{P}_i$. For example, $\mathcal{P}_i = (3, 4, 2)$ indicates that path $\mathcal{P}_i$ consists of the third channel at hop 1, the fourth channel at hop 2, and the second channel at hop 3. Once channel pairing is determined at all the relays, the total $N$ disjoint paths $\mathcal{P}_1, \cdots, \mathcal{P}_N$ can be identified from the source to the destination.

Denote the power allocated to channel $n$ over hop $m$ by $P_{m,n}$. The SNR obtained on this channel is represented by $\gamma_{m,n} = P_{m,n}a_{m,n}$. For each path $\mathcal{P}_i$, let $\tilde{\gamma}_{m,i} \triangleq \gamma_{m,c(m,i)}$ represent the SNR seen over hop $m$ on this path.

Let $\mathbf{P}_i = (P_{1,c(1,i)}, \cdots, P_{M,c(M,i)})$ be the PA vector for all channels along path $\mathcal{P}_i$. The source-destination equivalent SNR of path $\mathcal{P}_i$ is denoted by $\gamma_{SD}(\mathcal{P}_i, \mathbf{P}_i)$. For AF relaying, it is given by [83],

$$\gamma_{SD}^{AF}(\mathcal{P}_i, \mathbf{P}_i) = \left( \prod_{m=1}^{M} \left(1 + \frac{1}{\tilde{\gamma}_{m,i}} \right) - 1 \right)^{-1},$$

and, in Section 3.3.3, we will also use its upper bound [83],

$$\gamma_{SD}^{AF}(\mathcal{P}_i, \mathbf{P}_i) \approx \left( \sum_{m=1}^{M} \frac{1}{\tilde{\gamma}_{m,i}} \right)^{-1},$$

whose approximation gap vanishes as the SNR becomes large. For DF relaying, we have

$$\gamma_{SD}^{DF}(\mathcal{P}_i, \mathbf{P}_i) = \min_{m=1,\cdots,M} \tilde{\gamma}_{m,i}.$$  \hspace{1cm} (3.3)

We consider two types of power constraints:

**Total power constraint:** The power assignment $P_{m,n}$, for $m = 1 \cdots M$ and $n = 1 \cdots N$, must satisfy the following aggregated power constraint

$$\sum_{m=1}^{M} \sum_{n=1}^{N} P_{m,n} \leq P_t.$$  \hspace{1cm} (3.4)
Chapter 3. Jointly Optimal CP and PA for Multi-channel Multi-hop Relaying

Individual power constraint: The power assignment $P_{m,n}$, for $n = 1, \cdots, N$, needs to satisfy the power constraint of the individual node $m$, i.e.,

$$\sum_{n=1}^{N} P_{m,n} \leq P_{mt}, \quad m = 1, \cdots, M,$$  \hspace{1cm} (3.5)

where $P_{mt}$ denotes the maximum allowable power at node $m$.

3.1.1 Objective

Our goal is to design a jointly optimal CP and PA strategy to maximize the source-destination rate under multi-hop relaying. The source-destination rate achieved through path $P_i$ is given by

$$R_{SD}(P_i, P_i) = \frac{1}{F_s} \log_2(1 + \gamma_{SD}(P_i, P_i)),$$

where $F_s$ is the spatial reuse factor. In multi-hop relaying that allows concurrent transmissions, $F_s$ takes value between 2 and $M$ (since $F_s \geq 2$ under the half-duplex assumption). The sum rate of all paths determines the total source-destination rate of the system, denoted as $R_t$, i.e.,

$$R_t = \sum_{i=1}^{N} R_{SD}(P_i, P_i).$$ \hspace{1cm} (3.6)

It is a function of both $\{P_i\}$ and $\{P_i\}$, which should be jointly optimized:

$$\max_{\{P_i\}, \{P_i\}} R_t$$ \hspace{1cm} (3.7)

subject to (3.4) or (3.5),

$$P_i \succeq 0, \quad i = 1, \cdots, N,$$  \hspace{1cm} (3.8)

where $\succeq$ signifies element-wise inequality.
3.2 Optimal Multi-hop CP Under Fixed PA

In this section, we first consider the case when PA is fixed and given. In this case, the optimization problem in (3.7) can be re-written as

$$\max_{\{P_i\}} \sum_{i=1}^{N} R_{SD}(P_i, P_i),$$

and the optimal CP \(\{P^*_i\}\) is a function of \(\{P_i\}\). To simplify the notation, in this section we rewrite \(R_{SD}(P_i)\) and \(\gamma_{SD}(P_i)\) and drop their dependency on \(P_i\) with the understanding that \(\{P_i\}\) is fixed. In the following, we solve (3.9) to obtain the optimal CP scheme under this fixed PA. We emphasize that here the generalization from the dual-hop case to the multi-hop case is non-trivial. Intuitively, there is no obvious argument to decouple the sequence of pairings at all \((M-1)\) relays. Indeed, the equivalent incoming channel from a source to a relay and the equivalent outgoing channel from that relay to the destination depend on how the channels are paired over multiple hops. However, we will show that the optimal CP solution over multiple hops can in fact be decomposed into \((M-1)\) independent CP problems, where the mapping of incoming and outgoing channels at each relay is only based on the sorted SNR over those channels, and therefore can be performed individually per hop.

In the following, we first establish the optimality of the sorted-SNR CP scheme for the case of \(M = 3\) and \(N = 2\), and then we extend the result to arbitrary \(M\) and \(N\).

3.2.1 Optimal CP for Three-Hop Relaying

Two-channel case \((N = 2)\)

We first consider a three-hop relaying network with two channels, as depicted in Fig. 3.2. Without loss of generality, we assume channel 1 exhibits equal or larger SNR than channel 2 over all the three hops, i.e.,

$$\gamma_{m,1} \geq \gamma_{m,2}, \text{ for } m = 1, 2, 3. \tag{3.10}$$

The optimal CP scheme for this case is given in Proposition 3.1.
Proposition 3.1. For $M = 3$ and $N = 2$, the solution to (3.9) is the sorted-SNR CP scheme performed on each relay, i.e., $\{P_i^*\} = \{(1, 1, 1), (2, 2, 2)\}$ under condition (3.10).

Proof. The proof essentially examines possible path selections and shows that the sorted-SNR per relay provides the highest source-destination sum rate for both AF and DF relaying. See Appendix A 7.1.1 for details.

Multi-channel case ($N \geq 2$)

Here, we provide an argument to extend the result in Proposition 3.1 to a system with an arbitrary number of channels.

Proposition 3.2. For $M = 3$ and $N \geq 2$, the solution to (3.9) is the sorted-SNR CP scheme performed on each relay.

Proof. Suppose the optimal pairing does not follow the pairing rule of sorted SNR. There is at least one relay (say, Relay 2) that has two pairs of incoming and outgoing channels that are mismatched according to their SNR. That is, there exist two channels $i_1$ and $i_2$ over hop 2, and two channels $j_1$ and $j_2$ over hop 3 that are respectively paired with each other while $\gamma_{2,i_1} < \gamma_{2,i_2}$ and $\gamma_{3,j_1} > \gamma_{3,j_2}$. Note that these two channel pairs belong to two disjoint source-destination paths that can be regarded as a 2-channel relay system. From Proposition 3.1, we know that pairing channels $i_1$ with $j_2$ and $i_2$ with $j_1$ at relay 2 achieves a higher rate than the existing pairing over these two paths. Hence, by switching to this new pairing while keeping the other paths the same, we could increase the total rate. This contradicts our assumption on the optimality of
a non-sorted SNR CP scheme. Hence, there is no better scheme than sorted-SNR CP to obtain the maximum sum rate.

3.2.2 Optimal CP for Multi-hop Relaying

Building on Proposition 3.2, we next extend the result for 3-hop relaying to a relaying network with an arbitrary number of hops ($M \geq 3$) in the following proposition.

**Proposition 3.3.** The solution to (3.9) is the sorted-SNR CP scheme individually performed at each relay.

*Proof.* We prove by induction. It is shown in Proposition 3.2 that the sorted-SNR CP is optimal for $M = 3$. Suppose the claim holds for $M \leq L$. Now consider $M = L + 1$ as shown in Fig. 3.3(a). Let $\gamma_{eq,n}$ be the received SNR from the source to relay $L - 1$ over the $n$th incoming channel of that relay. We establish $N$ equivalent channels between the source and relay $L - 1$, with SNR over the $n$th channel as $\gamma_{eq,n}$. Then, the $(L + 1)$-hop relaying network can be converted to a 3-hop network, with an equivalent relay whose incoming channels have SNR $\{\gamma_{eq,n}\}$ and outgoing channels remain the same as those of relay $L - 1$, as shown in Fig. 3.3(b). Hence, from Proposition 3.2, the optimal CP is the one where $\{\gamma_{eq,n}\}$ and $\{\gamma_{L,n}\}$ are sorted and paired at this equivalent relay, and $\{\gamma_{L,n}\}$ and $\{\gamma_{L+1,n}\}$ are sorted and paired at relay $L$. Note that the sorted-SNR pairing at relay $L$ is independent of how the channels are paired at the other relays.

Next, ignore relay $L$ and replace it by equivalent channels from relay $L - 1$ to the destination. We now have a $L$-hop network. From the induction hypothesis, the sorted-SNR CP is optimal. In particular, the incoming and outgoing channels at each of relays $1, 2, \ldots, L - 2$ are sorted by their SNR and paired. Since the SNRs $\{\gamma_{eq,n}\}$ at the equivalent relay are computed by applying (3.1) or (3.3) over these *sorted and paired* channels from the source to relay $L - 1$, it is not difficult to see that $\{\gamma_{L-1,n}\}$ and $\{\gamma_{eq,n}\}$ are ordered in the same way. Therefore, sorting and pairing $\{\gamma_{eq,n}\}$ and $\{\gamma_{L,n}\}$ at the equivalent relay is the same as sorting and pairing
\{\gamma_{L-1,n}\} and \{\gamma_{L,n}\} at relay \(L - 1\). Thus, we conclude that at each of relay \(1, \cdots, L\), the incoming and outgoing channels are sorted and paired in order of their SNR.

The significance of Proposition 3.3 is that the optimal CP for \(M\)-hop relaying is decoupled into \((M - 1)\) individual pairing schemes at each relay, each solely based on the SNR of incoming and outgoing channels. This decoupling not only reduces the pairing complexity, but also reveals the distributed nature of optimal CP among multiple relays, thus allowing simple implementation that can easily adapt to network topology changes.

Remark: We point out that the existing result of optimal CP strategy for dual-hop relaying is not sufficient for the induction to prove Proposition 3.3. Notice that, in the proof, an \(M\)-hop network \((M > 3)\) was transformed into an equivalent 3-hop network. Reducing a 3-hop network to a dual-hop network would require combining relay nodes with either the source or the destination to form an equivalent node and equivalent channel gain. The dual-hop result can only be applied to pairing with the equivalent channels, but is not sufficient to show the actual physical channel should follow the same pairing strategy. Therefore, Proposition 3.2 is necessary as the basis to prove the general \(M\)-hop case.

In addition, in [61], the L-superadditivity property [84] is used to show that the sorted-SNR CP is optimal in dual-hop AF relaying for sum-rate maximization. That is, if the source-destination rate over each path can be shown to be L-superadditive, it follows that sorted-SNR pairing is optimal. However, L-superadditivity does not hold for the rate function in general multi-hop relaying, where the source-destination rate is a higher dimensional function defined on \(\mathbb{R}^M\) with respect to \(\gamma_{1,n}, \cdots, \gamma_{M,n}\), for a given \(n\). Thus, a similar proof for the optimality of sorted-SNR in the dual-hop case is not available to the general multi-hop case.

### 3.3 Jointly Optimal CP and PA: A Separation Principle

So far, given a fixed PA scheme, we have found that the optimal CP scheme for (3.9) is SNR based, which depends on the transmission power allocated to each channel. We next present
the solution for (3.7) by jointly optimizing CP and PA.

The apparent coupling of CP and PA makes a direct exhaustive search for the jointly optimal solution prohibitively complex. Instead, we will show that the joint optimization problem can be decoupled into two separate CP and PA subproblems. Specifically, we prove that the jointly optimal solution is obtained by pairing channels based on the order of their channel gains (normalized against the noise power), followed by optimal PA over the paired channels. This separation principle holds for a variety of scenarios, including AF and DF relaying under either total or individual power constraints.

Our argument for the separation principle is briefly summarized as follows. We first show that, at a global optimum, the channel with a higher channel gain exhibits a larger SNR. This relation reveals that the SNR-based ordering of channels is the same as the one based on channel gain at optimality. Hence, we conclude that the sorted CP scheme based on channel gain is optimal when PA is also optimized.
3.3.1 Ordering Equivalence at Optimality

Let $\gamma_{m,n}^*$ be the received SNR under the optimal PA solution $\{P_n^*\}$ for hop $m$ and channel $n$. For both total and individual power constraints, the following proposition establishes the equivalence between channel-gain ordering and SNR-based ordering at the optimality.

**Proposition 3.4.** In the optimal CP and PA solution for (3.7), at each hop, the channel with better channel gain also provides a higher received SNR, i.e., $a_{m,i} \geq a_{m,j}$ implies $\gamma_{m,i}^* \geq \gamma_{m,j}^*$, for $m = 1, \cdots, M$; $i, j \in \{1, \cdots, N\}$, and $i \neq j$.

**Proof.** We first provide a proof for a multi-hop system consisting of two channels. We then explain how it can be extended to a system with an arbitrary number of channels.

$N = 2$ We prove the proposition by contradiction. Let $P_1$ and $P_2$ represent the two disjoint source-destination paths corresponding to the optimal CP scheme. Consider any hop $m$ along these paths. Without loss of generality, let channel 1 belong to $P_2$, channel 2 belong to $P_1$, and $a_{m,1} \geq a_{m,2}$. Suppose at optimality $\gamma_{m,1}^* < \gamma_{m,2}^*$, i.e., $P_{m,2}^*a_{m,2} > P_{m,1}^*a_{m,1}$, where $P_{m,1}^*$ and $P_{m,2}^*$ are the power allocated to channels 1 and 2, respectively. Let $P_{mt} = P_{m,1}^* + P_{m,2}^*$.

Consider the following alternate allocation of power between channels 1 and 2 over hop $m$

$$P_{m,1} = \frac{a_{m,2}}{a_{m,1}}P_{m,2}^*, \quad P_{m,2} = \frac{a_{m,1}}{a_{m,2}}P_{m,1}^*.$$  

(3.11)

We further swap the two channels so that channel 1 belongs to path $P_1$ and channel 2 belongs to path $P_2$. Since $P_{m,1}a_{m,1} = P_{m,2}^*a_{m,2}$ and $P_{m,2}a_{m,2} = P_{m,1}^*a_{m,1}$, the above procedure of power re-allocation and channel swapping does not change the end-to-end rate.

$$P_{m,1} + P_{m,2} = \frac{a_{m,2}}{a_{m,1}}(P_{mt} - P_{m,1}^*) + \frac{a_{m,1}}{a_{m,2}}P_{m,1}^*$$

$$= \frac{a_{m,2}}{a_{m,1}}P_{mt} + \frac{(a_{m,1})^2 - (a_{m,2})^2}{a_{m,1}a_{m,2}}P_{m,1}^*$$

$$< \frac{a_{m,2}}{a_{m,1}}P_{mt} + \frac{(a_{m,1})^2 - (a_{m,2})^2}{a_{m,1}a_{m,2}} \frac{a_{m,2}}{a_{m,1} + a_{m,2}}P_{mt}$$

$$= P_{mt},$$  

(3.12)
where inequality (3.12) is obtained from our assumption that $P_{m,1}a_{m,1} < P_{m,2}a_{m,2}$, which can be rewritten as $P_{m,1} < \frac{a_{m,2}}{a_{m,1}+a_{m,2}}P_{mt}$, and that $a_{m,1} \geq a_{m,2}$. This contradicts our initial assumption that the original PA is globally optimal.

\[ N > 2 \]

A similar proof by contradiction as it was used in Section 3.2.1 for Proposition 3.2 can be applied to generalize the above result to $N > 2$. For an $N$-channel relay system with $N > 2$, suppose the optimal CP scheme follows the pairing rule of the sorted CP based only on SNR gain and not channel gain. As a result, there is at least one hop over which, between two channels, the channel with better channel gain demonstrates a lower SNR. These two channels essentially belong to the two source-destination paths that can be considered as a 2-channel relay system. From the above, we know that by just swapping these two channels and applying the alternate allocation of power in (3.11), the sum power is reduced while maintaining the same rate. This leads to a contradiction of our early assumption on the optimality of the sorted CP not being conducted based on the channel gain.

\[ \square \]

### 3.3.2 Separation Principle

**Proposition 3.5.** The joint optimization of CP and PA in (3.7) can be separated into the following two steps:

1. Obtain the optimal CP $\{P^*_i\}$. The optimal CP $\{P^*_i\}$ is independent of $\{P^*_i\}$ and is performed individually at each relay in the order of sorted channel gain.

2. Obtain the optimal PA $\{P^*_i\}$ under the optimal CP $\{P^*_i\}$:

\[
\{P^*_i\} = \arg\max_{\{P_i\}} \sum_{i=1}^{N} R_{SD}(P^*_i, P_i) \quad \text{subject to (3.4) or (3.5).} \tag{3.13}
\]

**Proof.** From Proposition 3.3, with optimal PA $\{P^*_i\}$, the sorted-SNR CP gives the optimal $\{P^*_i\}$. From Proposition 3.4, at optimality, the sorted-SNR CP is equivalent to sorting channel gains, which does not require the knowledge of $\{P^*_i\}$. The optimal $\{P^*_i\}$ then can be obtained under the optimal CP, and we have the separation principle. \[ \square \]
Decoupling the CP strategy from PA strategy significantly reduces the problem search space. In addition, the optimal CP strategy in the presence of multiple hops is further decoupled into independent sorting problems at each hop, which only depends on the channel gain on the incoming and outgoing channels. The complexity of the optimal CP strategy for each hop is that of sorting channel gain, which is $O(N \log N)$. Therefore, the total complexity of the joint CP and PA optimization is $O(MN \log N)$ in addition to the complexity of PA optimization.

### 3.3.3 Optimal PA for Multi-hop Relaying

So far we have obtained the optimal CP at all relays. We next find the optimal PA solution for a given CP scheme as in (3.13). With the channels paired at each relay, the system can be viewed as a regular multi-hop system. Without loss of generality, we assume the channel gains at each hop are in descending order according to their channel index, i.e., $a_{m,1} \geq a_{m,2} \geq \cdots \geq a_{m,N}$, for $m = 1, \cdots, M$. From Proposition 3.5, the channels with the same index are paired, and a path with the optimal CP consists of all the same channel index, i.e., $P_i^* = (i, \cdots, i)$. In the following, we consider the PA optimization problem for total power and individual power constraints separately.

#### Total Power Constraint

The optimal PA solution with a total power constraint for a multi-hop relaying OFDM system was obtained in [68]. The results can be directly applied here. We briefly state the solution for completeness.

The PA optimization problem in (3.13) with a total power constraint has the classical water-filling solution

$$ P_i^* = \left[ \frac{1}{\lambda} - \frac{1}{a_{eq,i}} \right]^+ \quad \text{for } i = 1, \cdots, N, \quad (3.14) $$

where $[x]^+ = \max(x, 0)$. The Lagrange multiplier $\lambda$ is chosen such that the power constraint in (3.4) is met, and $a_{eq,i}$ is an equivalent channel gain over the path $P_i^*$. 
For DF relaying, the equivalent channel gain, denoted as $a_{eq,i}^{DF}$, is given by

$$a_{eq,i}^{DF} = \left( \sum_{m=1}^{M} \frac{1}{a_{m,i}} \right)^{-1}, \quad i = 1, \cdots, N.$$  \hspace{1cm} (3.15)

In other words, the equivalent channel gain is $N$ times the harmonic mean of the channel gain over each hop. It is obtained following the fact that, to maximize the source-destination rate on one path, the total power allocated to the path must be shared among the channels on this path such that all channels exhibit the same SNR. The power allocated to each transmitting node on path $P_i^*$ is given by

$$P_{m,i}^* = \frac{P_i^*}{a_{m,i} \sum_{m'=1}^{M} \frac{1}{a_{m',i}}}.$$  \hspace{1cm} (3.16)

For AF relaying, the exact expression for equivalent channel gain on path $P_i^*$ is difficult to obtain. However, its upper bound approximation can be expressed as

$$a_{eq,i}^{AF} \approx \left( \sum_{m=1}^{M} \frac{1}{\sqrt{a_{m,i}}} \right)^{-2}, \quad i = 1, \cdots, N.$$  \hspace{1cm} (3.17)

In this case, the equivalent channel amplitude (normalized against noise standard deviation) is $N$ times the harmonic mean of the channel amplitude over each hop. It is obtained using the upper bound approximation of equivalent SNR in (3.2) over a path. The power allocated to each transmitting node on path $P_i^*$ is given by

$$P_{m,i}^* = \frac{P_i^*}{\sqrt{a_{m,i}} \sum_{m'=1}^{M} \frac{1}{\sqrt{a_{m',i}}}}.$$  \hspace{1cm} (3.18)

The PA solution in (3.14) requires global channel gain information and therefore needs to be implemented in a centralized fashion.

**Individual Power Constraint**

For DF relaying, the source-destination sum rate in (3.6) reduces to

$$R_t^{DF} = \frac{1}{P_s} \sum_{n=1}^{N} \min_{m=1, \cdots, M} \log_2(1 + P_{m,n} a_{m,n}).$$  \hspace{1cm} (3.19)
Maximizing (3.19) over \( \{P_{m,n}\} \) under individual power constraints in (3.5) can be cast into the following optimization problem using a set of auxiliary variables \( r = [r_1, \ldots, r_N]^T \):

\[
\max_{r, P} \frac{1}{F_s} \sum_{n=1}^{N} r_n
\]

subject to

1. \( r_n \leq \log_2(1 + P_{m,n}a_{m,n}), \quad m = 1, \ldots, M, \quad n = 1, \ldots, N; \)

2. \( \sum_{n=1}^{N} P_{m,n} \leq P_{mt}, \quad m = 1, \ldots, M; \)

3. \( P \geq 0, \)

where \( P \triangleq [P_{m,n}]_{M \times N} \). Since the objective function is linear, and all the constraints are convex, the optimization problem in (3.20) is convex. For such a problem, Slater’s condition holds [33], and the duality gap is zero. Thus, (3.20) can be solved in the Lagrangian dual domain.

Since the spatial reuse factor \( F_s \) is a constant, we drop it for simplicity without affecting the optimization problem. Consider the Lagrange function for (3.20),

\[
\mathcal{L}(P, r, \mu, \lambda) = \sum_{n=1}^{N} r_n - \sum_{n=1}^{N} \sum_{m=1}^{M} \mu_{m,n} (r_n - \log_2(1 + P_{m,n}a_{m,n})) - \sum_{m=1}^{M} \lambda_m \left( \sum_{n=1}^{N} P_{m,n} - P_{mt} \right)
\]

where \( \mu \triangleq [\mu_{m,n}]_{M \times N} \) with \( \mu_{m,n} \) being the Lagrange multiplier corresponding to constraint (i) in (3.20), and \( \lambda = [\lambda_1, \ldots, \lambda_M]^T \) with \( \lambda_m \) being the Lagrange multiplier associated with power constraint in (iii) in (3.20). The dual function is given by

\[
g(\lambda, \mu) = \max_{r, P} \mathcal{L}(P, r, \mu, \lambda)
\]

subject to \( P \geq 0. \)

Optimizing (3.21) over \( r \) for given \( P, \mu \) and \( \lambda \) yields

\[
\sum_{m=1}^{M} \mu_{m,n} = 1, \quad \text{for} \quad n = 1, \ldots, N.
\]

Substituting this into \( \mathcal{L}(P, r, \mu, \lambda) \), we obtain

\[
\mathcal{L}(P, r, \mu, \lambda) = \sum_{n=1}^{N} \sum_{m=1}^{M} (\mu_{m,n} \log_2(1 + P_{m,n}a_{m,n}) - \lambda_m P_{m,n}) + \sum_{m=1}^{M} \lambda_m P_{mt}.
\]
It is clear that the dual function \( g(\mu, \lambda) \) obtained by maximizing (3.24) can be decomposed into \( NM \) subproblems

\[
g(\mu, \lambda) = \sum_{n=1}^{N} \sum_{m=1}^{M} g_{mn}(\mu_{mn}, \lambda_m) + \sum_{m=1}^{M} \lambda_m P_{mt},
\]

with

\[
g_{mn}(\mu_{mn}, \lambda_m) = \max_{P_{mn}} L_{mn}(P_{mn}, \mu_{mn}, \lambda_m)
\]

subject to \( P_{mn} \geq 0 \)

where

\[
L_{mn}(P_{mn}, \mu_{mn}, \lambda_m) = \mu_{mn} \log_2 (1 + P_{mn}a_{mn}) - \lambda_m P_{mn},
\]

for \( m = 1, \ldots, M; n = 1, \ldots, N \). By applying KKT conditions [33] to (3.25), the optimal power allocation \( P_{mn}^* \), as a function of \( \mu_{mn} \) and \( \lambda_m \), is derived as

\[
P_{mn}^* = \left[ \frac{\mu_{mn}}{\lambda_m} - \frac{1}{a_{mn}} \right]^+, \tag{3.26}
\]

for \( n = 1, \ldots, N \) and \( m = 1, \ldots, M \), where \( \lambda_m \) is chosen to meet the power constraint in (3.20).

Finally, the optimization problem in (3.20) is equivalent to the dual problem

\[
\min_{\mu, \lambda} g(\mu, \lambda)
\]

subject to \( \mu \succeq 0, \lambda \succeq 0; \)

\[
\sum_{m=1}^{M} \mu_{mn} = 1, \quad \text{for} \quad n = 1, \ldots, N.
\]

This dual problem can be efficiently solved by using the *projected* subgradient method [85]. Analogous to a common subgradient method, a sequence of Lagrange multipliers is generated which converges to the optimal \( \lambda^* \) and \( \mu^* \) minimizing \( g(\mu, \lambda) \). As stated in Section 2.2.3, the convergence is achieved provided that a suitable step size is chosen at each iteration [85]. The difference between projected and normal subgradient methods lies in having an extra constraint
\[ \sum_{m=1}^{M} \mu_{m,n} = 1. \]

To satisfy this constraint, at each iteration, the projected subgradient method projects the columns of \( \mu \) (obtained by subgradient method) onto a unit space to attain a set of feasible multipliers. At each iteration, a subgradient of \( g(\mu, \lambda) \) at the current values of \( \mu_{m,n} \) and \( \lambda_m \) is required. Let \( [\theta \mu, \theta \lambda]^T \) denote the subgradient, where \( \theta \mu = [\theta_{\mu_{1,1}}, \cdots, \theta_{\mu_{M,N}}]^T \) and \( \theta \lambda = [\theta_{\lambda_{1}}, \cdots, \theta_{\lambda_{M}}]^T \). It is obtained from (3.24) as

\[ \theta_{\mu_{m,n}} = \log_2 \left(1 + P_{m,n}^* a_{m,n}\right), \tag{3.28} \]

for \( m = 1, \cdots, M \) and \( n = 1, \cdots, N \), and

\[ \theta_{\lambda_{m}} = P_{m,n}^* - \sum_{n=1}^{N} P_{m,n}^*, \]

for \( m = 1, \cdots, M \), where \( P_{m,n}^* \) is obtained from (3.26).

For completeness, we summarize the projected subgradient algorithm for solving the dual problem:

1. Initialize \( \lambda^{(0)} \) and \( \mu^{(0)} \).

2. Given \( \lambda^{(l)}_m \) and \( \mu^{(l)}_{m,n} \), obtain the optimal values of \( P_{m,n}^* \) in (3.26) for all \( m \) and \( n \).

3. Update \( \lambda^{(l)} \) through

\[ \lambda^{(l+1)}_m = \left[\lambda^{(l)}_m - \theta_{\lambda_{m}} \nu_{\lambda}\right]^+, \]

for \( m = 1, \cdots, M \). Similarly, update \( \mu^{(l)}_{m,n} \) followed by unitary space projection i.e.,

\[ \mu^{(l+1)}_{m,n} = \frac{\hat{\mu}^{(l+1)}_{m,n}}{\sum_{j=1}^{M} \hat{\mu}^{(l+1)}_{j,n}}, \tag{3.29} \]

where

\[ \hat{\mu}^{(l+1)}_{m,n} = \left[\mu^{(l)}_{m,n} - \theta_{\mu_{m,n}} \nu_{\mu}\right]^+, \tag{3.30} \]

for \( m = 1, \cdots, M \); and \( n = 1, \cdots, N \). \( \nu_{\lambda}^{(l)} \) and \( \nu_{\mu}^{(l)} \) are the step sizes at the \( l \)th iteration for multipliers \( \mu \) and \( \lambda \), respectively.

4. Let \( l = l + 1 \); repeat from Step 2 until convergence.
With the optimal $\lambda^*$ and $\mu^*$, the optimal power solution $P^*$ is determined as in (3.26), i.e., for $m = 1, \cdots, M$ and $n = 1, \cdots, N$,

$$P^*_{m,n} = \left[ \frac{\mu^*_{m,n}}{\alpha \lambda^*_m} - \frac{1}{a_{m,n}} \right]^+, \quad (3.31)$$

where $\mu^*_{m,n}$ satisfies the constraint (3.23), and at the same time, $\mu^*_{m,n}$ and $\lambda^*_m$ are chosen so that the individual power constraints in constraint (ii) of (3.20) are met.

The expression of $P^*_{m,n}$ in (3.31) provides some insight on the structure of the optimal PA for multi-hop DF relaying: For a given $\mu$, the power allocation across channels at each node is individually determined following a scaled version of the water-filling approach based on the channel gain. The scales are determined jointly among different hops to satisfy the condition of $\mu$ in (3.23). It essentially requires the received SNR $\gamma_{m,n}$ at each hop of the same path to be equal.

**AF Relaying:** We now consider the PA problem for AF relaying. Unlike DF, the achievable source-destination sum rate for AF is not generally concave in $\{P_{m,n}\}$. Therefore, we have a non-convex optimization problem formulated as

$$\max_{P} \frac{1}{F_s} \sum_{n=1}^{N} \log_2 \left( 1 + \left( \prod_{m=1}^{M} \left( 1 + \frac{1}{P_{m,n} a_{m,n}} \right) - 1 \right)^{-1} \right) \quad (3.32)$$

subject to

- $i)$ $\sum_{n=1}^{N} P_{m,n} \leq P_{mt}, \quad m = 1, \cdots, M$

- $ii)$ $P \succeq 0$.

To find the PA solution we resort to an upper bound of the sum rate in (3.32). Based on (3.2), an upper-bound approximation for the $M$-hop source-destination sum rate is given by

$$R_{t}^{up} = \frac{1}{F_s} \sum_{n=1}^{N} \log_2 \left( 1 + \left( \sum_{m=1}^{M} \frac{1}{P_{m,n} a_{m,n}} \right)^{-1} \right). \quad (3.33)$$

This upper bound becomes tight as the channel gain $a_{m,n}$ over each channel increases. Therefore, the PA solution obtained using (3.33) is asymptotically optimal.

**Lemma 3.1.** $R_{t}^{up}$ in (3.33) is concave with respect to $\{P_{m,n}\}$. 
Proof. The proof follows from the concavity of (3.2) with respect to $\{P_{m,n}\}$, which can be shown by considering its Hessian matrix. The details are given in Appendix A 7.1.2. □

Given Lemma 3.1, the optimization of $\{P_{m,n}\}$ to maximize $R_{\text{up}}^\text{opt}$ is a convex optimization problem, and we can solve it in the Lagrangian dual domain using KKT conditions [33]. Although a closed-form or semi-closed-form solution for $\{P_{m,n}\}$ is difficult to obtain in this case, we can solve it numerically via standard convex optimization tools.

### 3.3.4 Extension to Multiple-Destination Case

The system model we consider so far assumes a single pair of source and destination communicating through $M$-hop relaying. In this section, we show that the results in the previous sections can be extended to a multi-hop multi-destination relaying network for sum-rate maximization.

Specifically, we consider a single source node communicating with $K$ users through $M$-hop relaying via $(M-1)$ common relay nodes. In this multi-destination system, the last relay conducts CP as well as channel-user assignment. In channel-user assignment, the relay partitions the $N$ outgoing channels into $K$ subsets, assigning one for each user for data forwarding. To maximize the end-to-end sum-rate of $K$ users, the joint optimization now involves CP, channel-user assignment, and PA. Despite the correlation between CP and channel-user assignment, we show that the results for the single source-destination case can be directly extended to the multi-destination case. To see this, we notice that all users share the common $(M-1)$-hop relay channels and the channel-user assignment is performed at the last relay for the $M$th hop. Given a channel-user assignment, this multi-destination system can be viewed equivalently as a single-destination system, and the results of optimal CP and PA in the previous sections apply. To optimize the channel-user assignment, it is not difficult to see that, for any given pairing in the first $(M-1)$ hops, at the last relay, assigning each outgoing channel to a user who has

---

While this model is appropriate to a downlink scenario, the result and proposed solution is applicable to the uplink scenario by swapping the role of source and destination nodes, where the given power constraint concerns the source nodes only in terms of their total transmission power.
the strongest channel gain among all \( K \) users maximizes the end-to-end rate. This result is summarized in the following proposition.

**Proposition 3.6.** For a \( M \)-hop \( K \)-user relaying network described above, the jointly optimal solution of CP, channel-user assignment, and PA that maximizes the end-to-end sum-rate is obtained via two steps:

1. **Channel-user assignment:** In the last hop, the \((M - 1)\)th relay assigns its \( i \)th outgoing channel to a user exhibiting the strongest channel gain among \( K \) users over that channel;

2. **CP and PA:** Under the channel-user assignment from step 1, apply the jointly optimal CP and PA solution of the single source-destination case.

We note that the above result for the multi-destination scenario holds for both DF and AF relaying, and for total and individual power constraints.

### 3.4 Numerical Results

In this section, we provide simulation examples to evaluate and compare the performance of the optimal joint CP and PA scheme with that of suboptimal CP and PA alternatives. We study different factors that affect the performance gap under these schemes.

Besides the jointly optimized CP and PA scheme, the following suboptimal schemes are used for comparison: 1) **Uniform PA with CP:** the optimal sorted channel gain based CP is first performed. At each transmitting node, the power is uniformly allocated on each subcarrier. In addition, for total power constraint, the total power is also uniformly allocated to each transmitting node. Therefore, for individual power constraint, \( P_{m,n} = \frac{P_t}{N} \); and for total power constraint, \( P_{m,n} = \frac{P_t}{MN} \); 2) **Opt. PA without CP:** only power allocation is optimized but no pairing, i.e., the same incoming and outgoing channels are assumed; 3) **Uniform PA without CP:** the same incoming and outgoing channels are assumed, then uniform PA as in the case 1 is used.
CHAPTER 3. JOINTLY OPTIMAL CP AND PA FOR MULTI-CHANNEL MULTI-HOP RELAYING

We use an OFDM system as an example of a multi-channel system, and refer each subcarrier as a channel in this case. For the multi-hop setup, equal distance is assumed from hop to hop, and is denoted by $d_r$. No direct link between source and destination is available. The spatial reuse factor is set to $F_s = 3$ (i.e., interference is assumed negligible three hops away). We assume $M = 4$, unless it is otherwise specified. An $L$-tap frequency-selective fading channel is assumed for each hop. We define the average SNR as the average received SNR over each subcarrier at each receiving node under uniform power allocation. Specifically, it is defined for different power constraint as follows: under the total power constraint, \( \text{SNR}_{\text{avg}} \triangleq \frac{P_d r^{-\alpha}}{M N \sigma^2} \), where $\alpha$ denotes the pathloss exponent and $\sigma^2$ the noise variance; under the individual power constraint, \( \text{SNR}_{\text{avg}} \triangleq \frac{P_{\text{mt}} d^{-\alpha}}{N \sigma^2} \).

Impact of the average SNR We compare the performance of various CP and PA schemes at different average SNR levels. Fig. 3.4 shows the normalized source-to-destination per-subcarrier rate vs. the average SNR defined above, for DF relaying under the total power constraint. The number of channels is set to $N = 64$. We observe that joint optimization of CP and PA provides significant performance improvement over the other schemes. In particular, compared with uniform PA without CP, the optimal CP alone provides 4dB gain, and subsequently optimally allocating power provides an additional 1.5-2dB gain. Interestingly, it is evident that channel pairing alone provides more performance gain than power allocation alone does.

Fig. 3.5 plots the normalized source-to-destination per-subcarrier rate vs. the average SNR for AF relaying under the total power constraint. Again, $N = 64$ is used. For schemes with PA optimization, the upper-bound $R^{\text{up}}_t$ in (3.33) is used to obtain the PA solution. The actual rate $R_t$ obtained (as in the objective function in (3.32)) with such PA solution provides a lower bound on the rate under the optimal PA. In Fig. 3.5, for the jointly optimal CP and PA scheme and optimal PA without CP scheme, we plot both upper bound and lower bound of the rate for the optimal PA solution. We see that these two bounds become tighter as the average SNR increases, due to the improving accuracy of approximation $R^{\text{up}}_t$. The PA solution derived using
$R_{\text{up}}^\text{opt}$ becomes near optimal.

Comparing the performance of different CP and PA schemes shown in Fig. 3.5, it is seen that similarly as in the DF case, joint optimization of CP and PA provides noticeable improvement over the other schemes. The gain mainly comes from choosing CP optimally, which provides around 2dB gain over no CP schemes. We further observe that, with optimal CP, the gap between optimal and uniform PA vanishes at higher SNR, indicating that uniform PA achieves the optimal performance at a moderately high SNR range (around 15dB). Interestingly, this is not the case for the schemes without CP. The intuition behind this is the following: At relatively high SNR, it is known that the water-filling PA solution in (3.14) approaches a uniform allocation. Thus, the total power is approximately equally distributed to different paths. The power on each path $P_i^*$ is then further assigned optimally to each channel on the path according to (3.18), which is typically not uniform. The exception is when each hop exhibits a similar channel gain. This is more likely to occur as a result of channel pairing, where channels with the same rank, more likely with similar strength, are paired with each other. Therefore, with CP, the optimal PA approaches to a uniform allocation at a faster rate with increasing SNR. This interesting observation suggests that, because of CP, at moderately high SNR, we are able to reduce the centralized PA solution to a simple uniform PA which requires no global channel information without losing much optimality. Note that the same argument is applicable to DF relaying, but the optimal PA approaches to a uniform allocation at a much slower rate than that for AF, which can be shown by comparing (3.16) and (3.18). The range of SNR values under consideration is too small to see the same effect in Fig. 3.4.

Under individual power constraints, the performance comparison of CP and PA schemes are given in Figs. 3.6 and 3.7 for DF and AF relaying, respectively. We assume $N = 16$. These figures further demonstrate the significant improvement by jointly optimizing CP and PA, where most of the gain comes from optimal CP. In addition, under AF relaying, we again

\footnote{Note that, for water-filling PA, as $\text{SNR} \rightarrow \infty$, it approaches to a uniform allocation in all schemes with or without CP. The difference is the rate at which PA approaches to a uniform allocation.}
Figure 3.4: Normalized rate vs. the average SNR for DF OFDM relaying with $M = 4$ and $N = 64$ under total power constraint.

Figure 3.5: Normalized rate vs. the average SNR for AF OFDM relaying with $M = 4$ and $N = 64$ under total power constraint.
CHAPTER 3. JOINTLY OPTIMAL CP AND PA FOR MULTI-CHANNEL MULTI-HOP RELAYING

Figure 3.6: Normalized rate vs. the average SNR for DF OFDM relaying with $M = 4$ and $N = 16$ under individual power constraint.

Figure 3.7: Normalized rate vs. the average SNR for AF OFDM relaying with $M = 4$ and $N = 16$ under individual power constraint.
observe a near-optimal performance by uniform PA with CP at moderately high SNR. This suggests that, under individual power constraints, the optimal PA is close to a uniform allocation at high SNR as well when CP is adopted. This potentially simplifies greatly the PA implementation to achieve the optimal performance.

Impact of the Variation of Channel Gain In this experiment, we show how the level of channel gain variation across $N$ channels affects the performance of various CP and PA schemes. Towards this goal, we increase the number of taps of the time-domain frequency-selective channel (i.e., the maximum delay of the frequency-selective channel). This increases the level of variation of the corresponding frequency response. Figs. 3.8 and 3.9 plot the normalized per-subcarrier rate vs. the number of taps of the frequency-selective channel for DF and AF relaying, respectively. The number of subcarriers is set to $N = 64$ and $\text{SNR}_{\text{avg}} = 12\text{dB}$. As we see, the performance gap between the schemes with optimal CP and without CP increases as the level of channel gain variation increases. This demonstrates that the optimal CP schemes benefit from an increased level of channel diversity, which is utilized effectively through the channel pairing. On the other hand, the relative gain of optimal PA to uniform PA is insensitive to such change and remains constant.

Impact of the Number of Channels In this experiment, we examine the effect of the number of channels, under the same level of channel gain variation across channels, on the performance of various CP and PA schemes. For different $N$, the subcarrier spacing (i.e., bandwidth of each channel) is fixed. In order to set the same level of channel gain variation in frequency, we keep the maximum delay of the time-domain frequency-selective channel unchanged. Figs. 3.10 and 3.11 demonstrate the normalized per-subcarrier rate with respect to $N$ for DF and AF relaying, respectively. The average SNR is set to $\text{SNR}_{\text{avg}} = 12\text{dB}$. We observes that the gap between the two sets of schemes, with and without CP, widens as the number of channels increases. The reason behind this observation is that, as more channels becomes available, they can be exploited more judiciously for pairing, and therefore, more gain is achieved by CP. The different PA schemes are not sensitive to the change of $N$. 
Chapter 3. Jointly Optimal CP and PA for Multi-channel Multi-hop Relaying

Figure 3.8: Normalized rate vs. number of taps for DF OFDM relaying with $M = 4$, $N = 64$, and $SNR_{\text{avg}} = 12\text{dB}$ under total power constraint.

Figure 3.9: Normalized rate vs. number of taps for AF OFDM relaying with $M = 4$, $N = 64$, and $SNR_{\text{avg}} = 12\text{dB}$ under total power constraint.
Figure 3.10: Normalized rate vs. number of channels for DF OFDM relaying with $M = 4$ and $\text{SNR}_{\text{avg}} = 12\text{dB}$ under total power constraint.

Figure 3.11: Normalized rate vs. number of channels for AF OFDM relaying with $M = 4$ and $\text{SNR}_{\text{avg}} = 12\text{dB}$ under total power constraint.
Impact of the Number of Hops In this experiment, we study how the number of hops affects the performance of various CP and PA schemes. For this purpose, we increase the number of hops while keeping the distance between each hop unchanged. Figs. 3.12 and 3.13 illustrate the normalized per-subcarrier rate vs. the number of hops with total power constraint for DF and AF relaying, respectively. We set $N = 64$ and $\text{SNR}_{\text{avg}} = 12\text{dB}$. As expected, for all schemes, the normalized per-subcarrier rate decreases as the number of hops increases. For DF, this is because, on average, the minimum rate among all hops decreases as the number of hops increases; for AF, the rate decreases due to noise amplification over hops. Comparing different schemes, we observe that the performance of the jointly optimized CP and PA scheme has the slowest decay rate, and the performance of the schemes with CP decay is slower than those without CP. In other words, the gain of optimal CP and PA is more pronounced as the number of hops increases. A multiple-fold gain is observed at a higher number of hops.

![Figure 3.12: Normalized rate vs. number of hops for DF OFDM relaying with $N = 64$ and $\text{SNR}_{\text{avg}} = 12\text{dB}$ under total power constraint.](image-url)
3.5 Summary

This chapter has focused on the problem of jointly optimizing spectrum and power allocation to maximize the source-to-destination sum rate for a multi-channel $M$-hop relaying network. It is shown that the joint CP and PA problem can be decoupled into two separate CP optimization, and PA optimization subproblems (separation rule). This rule that holds for both AF and DF relaying strategies and single and multiple destinations significantly reduces the search space. It follows that the CP problem can be again decomposed into independent pairing problems at each relay based on sorted channel gains, independent of PA on each channel. The resulting decomposition allows simple distributed relay implementation for optimal operation, as well as easily adapting to the network topology changes. The solution for optimizing PA for both DF and AF relaying is also provided through a dual-decomposition approach.
Chapter 4

Jointly Optimal Channel Assignment and PA for Dual-Hop Multi-channel Multi-user Relaying

In this chapter, we consider the problem of jointly optimizing channel pairing, channel-user assignment, and power allocation, to maximize the weighted sum-rate, in a single-relay cooperative system with multiple channels and multiple users. Common relaying strategies are considered and transmission power constraints are imposed on both individual transmitters and the aggregate over all transmitters. This joint optimization problem naturally leads to a mixed-integer program. Despite the general expectation that such problems are intractable, we construct an efficient algorithm to find an optimal solution, which incurs computational complexity that is polynomial in the number of channels and the number of users. We further demonstrate through numerical experiments that the jointly optimal solution can significantly improve system performance over its suboptimal alternatives.

The rest of the chapter is organized as follows. In Section 4.1, we discuss the system model and formulate the joint optimization problem. For weighted sum-rate maximization
with DF relaying, we describe our framework of finding the optimal solution with polynomial complexity in Section 4.2. Extension to other relaying strategies is explained in Section 4.3. Numerical studies are presented in Section 4.4.

4.1 System Model and Problem Statement

We consider the multi-channel relaying scenario where a source communicates with \( K \) users via a single relay over \( N \) equal bandwidth channels, as illustrated in Fig. 4.1. We focus on the downlink in our analysis in this chapter, but the proposed solution framework can be adopted for the uplink by swapping the roles of the source and the users.

We denote by \( h_{sr}^i \), \( h_{rk}^i \), and \( h_{sk}^i \) the state of channel \( i \), for \( 1 \leq i \leq N \), over the first hop between the source and the relay, over the second hop between the relay and user \( k \), and over the direct link between the source and user \( k \), respectively. The additive noise on a channel at the relay and user \( k \) are modeled as i.i.d. zero-mean Gaussian random variables with variances \( \sigma_r^2 \) and \( \sigma_k^2 \), respectively. The channel state is assumed to be available at both the source and the relay, which enables them to dynamically assign channels and allocate power according to channel conditions.

Channel Assignment The relay transmits a processed version of the incoming data to its intended user using a specific relay strategy. The relay also conducts channel pairing (CP),
as described in chapter 2, and channel-user assignment. Through channel-user assignment, a subset of incoming-outgoing channels is assigned to each user. Clearly, channel-pairing choices are closely connected with how the channels are assigned to the users. We term the joint decision on channel pairing and channel-user assignment the channel assignment problem. As the different channels exhibit various quality, judicious channel assignment can potentially lead to significant improvement in spectral efficiency.

We say a path $P(m, n, k)$ is selected, if first-hop channel $m$ is paired with second-hop channel $n$, and the pair of channels $(m, n)$ is assigned to user $k$. We define indicator functions $\phi_{mnk}$ for channel assignment as follows:

$$\phi_{mnk} = \begin{cases} 
1, & \text{if } P(m, n, k) \text{ is selected}, \\
0, & \text{otherwise}.
\end{cases}$$  \hspace{1cm} (4.1)

We require that each channel pair be assigned to only one user, but a user may be assigned multiple channel pairs. Together with the one-to-one mapping assumption between first-hop channels and second-hop, $\phi_{mnk}$ is then constrained by

$$\sum_{n=1}^{N} \sum_{k=1}^{K} \phi_{mnk} = 1, \quad \forall m, \hspace{1cm} (4.2)$$

$$\sum_{m=1}^{N} \sum_{k=1}^{K} \phi_{mnk} = 1, \quad \forall n. \hspace{1cm} (4.3)$$

**Power Allocation (PA)** Along any path $P(m, n, k)$, the source and relay transmission powers are denoted by $P_{mnk}$ and $P_{mnk}^r$, respectively. We consider both individual power constraints,

$$\sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{k=1}^{K} P_{mnk}^s \leq P_s, \hspace{1cm} (4.4)$$

$$\sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{k=1}^{K} P_{mnk}^r \leq P_r, \hspace{1cm} (4.5)$$

and the total power constraint,

$$\sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{k=1}^{K} (P_{mnk}^s + P_{mnk}^r) \leq P_t, \hspace{1cm} (4.6)$$
where $P_s$, $P_r$, and $P_t$ are the maximum allowed transmission power by the source, the relay, and the combined source and relay, respectively. Each constraint above is either inactive (i.e., at optimality it is satisfied with strict inequality) or active (i.e., at optimality it is satisfied with equality). We consider the case where all active constraints are strictly active, i.e., if the problem is modified by changing the power limits by small amounts, at optimality the constraints remain active. This is without loss of generality, since any constraint that is active but not strictly active can be made inactive, by increasing the power limit by a small amount, without altering the problem solution. Define $p_{mnk} = (P_{mnk}^s, P_{mnk}^r, P_{mnk}^s + P_{mnk}^r)$ and $\pi = (P_s, P_r, P_t)$.

**Relaying Strategy** We initially focus on DF relaying but will later show how the proposed method can be applied to other relaying schemes, such as AF and CF. We consider a general case where, apart from the relay path, the direct links are available between the source and users. In this case, the signals received from the relay path and the direct link can be combined to improve the decoding performance. The DF relay, after decoding the received message from each incoming channel, forwards a version of the decoded message on an outgoing channel (second hop) to the intended user. The intended user collects the received signals in both time slots, applies maximum ratio combining, and decodes the message.

Considering the conventional repetition-coding based DF relaying [28][4], as implied in Section 2.1.4, the maximum achievable source-destination rate on path $\mathcal{P}(m,n,k)$ is given by

$$ R(m,n,k) = \frac{1}{2} \min \{ \log(1 + a_m P_{mnk}^s), \log(1 + c_{mk} P_{mnk}^s + b_{nk} P_{mnk}^r) \} , \quad (4.7) $$

where $a_m = \frac{|h_{mr}|^2}{\sigma_r^2}$, $b_{nk} = \frac{|h_{nk}|^2}{\sigma_k^2}$, and $c_{mk} = \frac{|h_{mk}|^2}{\sigma_k^2}$ are normalized channel power gains against the noise variance at the relay and user $k$, and the base of logarithm is 2.
4.1.1 Optimization Objective

Various rate-utility functions can be used to meet different performance and fairness criteria. Here we focus on maximizing the weighted sum-rate which leads to throughput improvement. Denoting by $w_k$ the relative weight for user $k$, such that $\sum_{k=1}^{K} w_k = 1$, we formulate the problem of weighted sum-rate maximization as

$$\max_{\Phi, P^s, P^r} \sum_{k=1}^{K} w_k \sum_{m=1}^{N} \sum_{n=1}^{N} \phi_{mnk} R(m, n, k)$$  \hspace{1cm} (4.8)

subject to

$$\phi_{mnk} \in \{0, 1\}, \forall m, n, k$$

$$P^s_{mnk} \geq 0, \ P^r_{mnk} \geq 0, \forall m, n, k,$$

where $\Phi \triangleq [\phi_{mnk}]_{N \times N \times K}$, $P^s \triangleq [P^s_{mnk}]_{N \times N \times K}$, and $P^r \triangleq [P^r_{mnk}]_{N \times N \times K}$. Given the relative weights and the channel gains on each path $\mathcal{P}(m, n, k)$, the optimization problem (4.8) finds the jointly optimal solution of channel pairing, channel-user assignment, and PA by optimizing $\Phi$, $P^s$, and $P^r$.

4.2 Weighted Sum-Rate Maximization for Multi-channel DF

The optimization in (4.8) is a mixed-integer programming problem, which in general has intractable complexity due to its combinatorial nature. However, in this section, we present a method to find an optimal solution with computational complexity growing only polynomial in the number of channels and users.

4.2.1 Convex Reformulation via Continuous Relaxation

The proposed approach is built on the reformulation of (4.8) into a convex optimization problem with a real-valued $\tilde{\Phi}$ and strong Lagrange duality. We later show that the reformulated problem is optimized by a binary $\Phi = \tilde{\Phi}$. 
We first substitute
\[ P_{mnk}^s = \frac{P_{mnk}^s}{\phi_{mnk}} \quad \text{and} \quad P_{mnk}^r = \frac{P_{mnk}^r}{\phi_{mnk}} \] (4.9)
into the objective of (4.8). This does not change the original optimization problem, since if \( \phi_{mnk} = 1 \), then (4.9) is trivially true; and if \( \phi_{mnk} = 0 \), then by l’Hôpital’s rule, \( \phi_{mnk} R(m, n, k) \) remains zero before and after the substitution. Indeed, it obviously preserves the optimality of PA to enforce \( P_{mnk}^s = P_{mnk}^r = 0 \) for all \((m, n, k)\) such that \( \phi_{mnk} = 0 \).

We then relax the binary constraint on \( \Phi \) by defining a continuous version of \( \phi_{mnk} \), denoted by \( \tilde{\phi}_{mnk} \), which may take any value in the interval \([0, 1]\). Then, the reformulated version of the optimization problem (4.8) can be written as
\[
\max_{\tilde{\Phi}, P^s, P^r} \sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{k=1}^{K} w_k \tilde{\phi}_{mnk} \min \left\{ \log(1 + a_m \frac{P_{mnk}^s}{\tilde{\phi}_{mnk}}), \log(1 + c_m \frac{P_{mnk}^s}{\tilde{\phi}_{mnk}} + b_n \frac{P_{mnk}^r}{\tilde{\phi}_{mnk}}) \right\}
\] (4.10)

s.t. \[
\sum_{m=1}^{N} \sum_{k=1}^{K} \tilde{\phi}_{mnk} = 1, \quad \forall n, \quad \sum_{n=1}^{N} \sum_{k=1}^{K} \tilde{\phi}_{mnk} = 1, \quad \forall m, \] (4.11)
\[
0 \leq \tilde{\phi}_{mnk} \leq 1, \quad \forall m, n, k, \] (4.12)
\[
(4.4), (4.5), (4.6), \] (4.13)
\[
P_{mnk}^s \geq 0, \quad P_{mnk}^r \geq 0, \quad \forall m, n, k. \]

The objective function (4.10) is concave in \((\tilde{\Phi}, P^s, P^r)\), since \( \tilde{\phi}_{mnk} \log(1 + a_m \frac{P_{mnk}^s}{\tilde{\phi}_{mnk}}) \) and \( \tilde{\phi}_{mnk} \log(1 + c_m \frac{P_{mnk}^s}{\tilde{\phi}_{mnk}} + b_n \frac{P_{mnk}^r}{\tilde{\phi}_{mnk}}) \) are the perspectives of the concave functions \( \log(1 + a_m P_{mnk}^s) \) and \( \log(1 + c_m P_{mnk}^s + b_n P_{mnk}^r) \), respectively.\(^1\) It is also noted that the minimum of two concave functions is a concave function. Furthermore, since all the constraints are affine, and there are obvious feasible points, Slater’s condition is satisfied [33]. Hence, the convex optimization problem (4.10) has zero duality gap, suggesting that a globally optimal solution can be found in the Lagrange dual domain.

\(^1\)The perspective of function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is defined as \( g(x, t) = tf(x/t) \), with domain \( \{(x, t) | x/t \in \text{dom } f, t > 0\} \). The perspective operation preserves concavity [33]. Here we include \( \phi_{mnk} = 0 \) in the domain of the perspectives. It is easy to see that they remain concave.
Using continuous relaxation on integer programming problems is not a new technique [20]. However, doing so typically leads only to heuristics or approximations. Clearly, solving a maximization problem with relaxed constraints generally gives only an upper bound to the original problem. In particular, all global optima for (4.10) do not necessarily give a binary \( \tilde{\Phi} \), which is required for (4.8). However, we next show that, in the problem under consideration, indeed there always exists a globally optimal solution to (4.10) consisting of a binary \( \tilde{\Phi} \), and the proposed approach ensures that such an optimal solution is found in polynomial time.

### 4.2.2 PA via Maximization of Lagrange Function over \( P^s \) and \( P^r \)

Consider the Lagrange function for (4.10),

\[
L(\tilde{\Phi}, P^s, P^r, \lambda) = \sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{k=1}^{K} \frac{w_{nk}}{2} \tilde{\phi}_{mnk} \min\{\log(1 + a_m P^s_{mnk}), \log(1 + c_{mk} P^s_{mnk} + b_{nk} P^r_{mnk})\} \\
- (\lambda_s + \lambda_t) \sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{k=1}^{K} P^s_{mnk} - (\lambda_r + \lambda_t) \sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{k=1}^{K} P^r_{mnk} + \lambda \pi^{tr},
\]  

(4.14)

where \( \lambda = (\lambda_s, \lambda_r, \lambda_t) \) is the vector of Lagrange multipliers associated with the power constraints (4.4), (4.5), and (4.6). The dual function is therefore

\[
g(\lambda) = \max_{\tilde{\Phi}, P^s, P^r} L(\tilde{\Phi}, P^s, P^r, \lambda)
\]  

s.t. \((4.11), (4.12), (4.13)\).

The above maximization of the Lagrange function can be carried out by first optimizing the PA given fixed \( \tilde{\Phi} \). The KKT conditions suggest that the maximization of (4.15) over \( P^s \) and \( P^r \) can be decomposed into \( N \times N \times K \) independent subproblems to find the optimal \( P^s_{mnk} \) and \( P^r_{mnk} \):

\[
\max_{P^s_{mnk} > 0, P^r_{mnk} > 0} L_{mnk}(\tilde{\phi}_{mnk}, P^s_{mnk}, P^r_{mnk}, \lambda)
\]  

(4.16)
where

\[
\mathcal{L}_{mnk}(\tilde{\phi}_{mnk}, P_{mnk}^s, P_{mnk}^r, \lambda) = \frac{w_k}{2} \tilde{\phi}_{mnk} \min \{ \log(1 + a_m \frac{P_{mnk}^s}{\tilde{\phi}_{mnk}}), \log(1 + c_m \frac{P_{mnk}^r}{\tilde{\phi}_{mnk}}) \} - (\lambda_s + \lambda_t) P_{mnk}^s - (\lambda_r + \lambda_t) P_{mnk}^r ,
\]

(4.17)

is the part of \( \mathcal{L}(\tilde{\Phi}, P^s, P^r, \lambda) \) that concerns only the path \( P(m, n, k) \).

It can be shown that the solution to (4.16) has the following form. The derivation details are given in Appendix B 7.2.1. Note that, since \( P^s_{mnk}^* \) and \( P^r_{mnk}^* \) depend on \( \tilde{\phi}_{mnk} \) in an obvious way, we simply present them as functions of \( \lambda \) for the rest of this section.

\[
(P_{mnk}^s(\lambda), P_{mnk}^r(\lambda)) = \begin{cases} \left( \frac{w_k b_{nk}}{\alpha(\lambda_s + \lambda_t)} - \frac{1}{a_m} \right)^+ \tilde{\phi}_{mnk}, 0 \right), & \text{if } a_m \leq c_m \\ (P_{mnk}^s(1), P_{mnk}^r(1)), & \text{if } a_m > c_m \text{ and } \frac{c_m}{\lambda_s + \lambda_t} < \frac{b_{nk}}{\lambda_r + \lambda_t} \approx \arg \max \max_{P_{mnk}^s, P_{mnk}^r \in \{(P_{mnk}^s, P_{mnk}^r), (P_{mnk}^s, P_{mnk}^r)\}} \mathcal{L}_{mnk}(\tilde{\phi}_{mnk}, P_{mnk}^s, P_{mnk}^r, \lambda), & \text{otherwise} \\
\end{cases}
\]

(4.18)

where \( \alpha \triangleq 2 \ln 2, [x]^+ \triangleq \max \{ x, 0 \} \), and

\[
(P_{mnk}^s(1), P_{mnk}^r(1)) = \left( \frac{w_k b_{nk}}{\alpha(\lambda_s + \lambda_t) + (a_m - c_m)(\lambda_r + \lambda_t)} - \frac{1}{a_m} \right)^+ \tilde{\phi}_{mnk}, \frac{a_m - c_m}{b_{nk}} P_{mnk}^s \\
(P_{mnk}^s(2), P_{mnk}^r(2)) = \left( \frac{w_k}{\alpha(\lambda_s + \lambda_t)} - \frac{1}{c_m} \right)^+ \tilde{\phi}_{mnk}, 0 \right). \]

(4.19)

### 4.2.3 Channel Assignment via Maximization of Lagrange Function over \( \tilde{\Phi} \)

To maximize the Lagrange function over \( \tilde{\Phi} \), we insert (4.18) into (4.17) and define

\[
A_{mnk}(\lambda) = \frac{1}{\phi_{mnk}} \mathcal{L}_{mnk}(\tilde{\phi}_{mnk}, P_{mnk}^s(\lambda), P_{mnk}^r(\lambda), \lambda).
\]

(4.20)
Note that $A_{mnk}(\lambda)$ is independent of $\tilde{\phi}_{mnk}$ because of the multiplication form of (4.18) by $\tilde{\phi}_{mnk}$. Then, (4.15) is equivalent to the following optimization problem over $\tilde{\Phi}$:

$$g(\lambda) = \max_{\tilde{\Phi}} \sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{k=1}^{K} \tilde{\phi}_{mnk} A_{mnk}(\lambda) + \lambda_s P_s + \lambda_r P_r + \lambda_t P_t$$

(4.21)

s.t. (4.11), (4.12).

To proceed, we present the following lemma on the decomposition of $\tilde{\Phi}$.

**Lemma 4.1.** Any matrix $\tilde{\Phi} = [\tilde{\phi}_{mnk}]_{N \times N \times K}$ with $0 \leq \tilde{\phi}_{mnk} \leq 1$ and satisfying (4.11) can be decomposed into one matrix $X = [x_{mn}]_{N \times N}$ and $MN$ vectors $y_{mn} = [y_{mn}^k]_{1 \times K}$, such that

$$\tilde{\phi}_{mnk} = x_{mn} y_{mn}^k, \forall m, n, k,$$

with $0 \leq x_{mn} \leq 1$ and $0 \leq y_{mn}^k \leq 1$, satisfying $\sum_{m=1}^{N} x_{mn} = 1, \forall n$, and $\sum_{k=1}^{K} y_{mn}^k = 1, \forall m$. Furthermore, any such matrix $X$ and a set of vectors $y_{mn}$ uniquely determines a matrix $\tilde{\Phi}$ that satisfies (4.11).

**Proof.** Given any $\tilde{\Phi} = [\tilde{\phi}_{mnk}]_{N \times N \times K}$ with $0 \leq \tilde{\phi}_{mnk} \leq 1$ and satisfying (4.11), let $x_{mn} = \sum_{k=1}^{K} \tilde{\phi}_{mnk}$. From $\sum_{m=1}^{N} \sum_{k=1}^{K} \tilde{\phi}_{mnk} = 1$, we have $\sum_{m=1}^{N} x_{mn} = 1$. Similarly, we have $\sum_{n=1}^{N} x_{mn} = 1$. Hence $0 \leq x_{mn} \leq 1$. Then, $y_{mn}^k$ can be constructed as

$$y_{mn}^k = \begin{cases} \tilde{\phi}_{mnk} / x_{mn}, & x_{mn} > 0 \\ 1/K, & x_{mn} = 0 \end{cases}$$

(4.22)

Hence, $\sum_{k} y_{mn}^k = 1$ and $0 \leq y_{mn}^k \leq 1$. Note that $1/K$ above is arbitrarily chosen, and the mapping from $\tilde{\Phi}$ to $(X, y_{mn})$ is one-to-many.

Given $X$ and $y_{mn}$ with $0 \leq x_{mn} \leq 1$ and $0 \leq y_{mn}^k \leq 1$, satisfying $\sum_{m=1}^{N} x_{mn} = 1, \forall m$, $\sum_{m=1}^{N} x_{mn} = 1, \forall n$, and $\sum_{k=1}^{K} y_{mn}^k = 1, \forall m, n$, clearly $0 \leq \tilde{\phi}_{mnk} = x_{mn} y_{mn}^k \leq 1$, and it is easy to verify that (4.11) is satisfied. This establishes the equivalence of $\tilde{\Phi}$ and the proposed decomposition.

Note that, even though the above decomposition can also be applied to a binary $\Phi$, as a trivial special case of Lemma 4.1, we require the general form of this lemma to deal with continuous values in $\tilde{\Phi}$, $X$, and $y_{mn}$. In particular, it is clear from the proof of Lemma 4.1 that
the mapping from $\tilde{\Phi}$ to $(X, \{y^{mn}\})$ is one-to-many, which is quite different from the binary case.

Lemma 4.1 implies that any optimization over $(X, y^{mn})$ also optimizes $\tilde{\Phi}$ for the same objective. This allows us to replace, in problem (4.21), $\tilde{\phi}_{mnk}$ with $x_{mn}y_k^{mn}$. Furthermore, the constant terms can be dropped from (4.21). Hence, we can equivalently seek solutions to the following problem

$$\max_{X, (y^{mn})} \sum_{m=1}^{N} \sum_{n=1}^{N} x_{mn} \sum_{k=1}^{K} y_k^{mn} A_{mnk}(\lambda)$$

subject to

$$\sum_{n=1}^{N} x_{mn} = 1, \forall m,$$

$$\sum_{m=1}^{N} x_{mn} = 1, \forall n,$$

$$0 \leq x_{mn} \leq 1, \forall m, n,$$

$$\sum_{k=1}^{K} y_k^{mn} = 1, \forall m, n,$$

$$0 \leq y_k^{mn} \leq 1, \forall m, n, k.$$  

The following two-stage solution is sufficient. First, the inner-sum term is maximized over $y_k^{mn}$ for each $(m, n)$ pair, i.e.,

$$A'_{mn}(\lambda) = \max_{y_k^{mn}} \sum_{k=1}^{K} y_k^{mn} A_{mnk}(\lambda)$$

subject to (4.25).

An optimal solution to (4.26) is readily obtained as

$$y_{k}^{mn^*} = \begin{cases} 
1, & \text{if } k = \arg\max_{1 \leq l \leq K} A_{mnl}(\lambda) \\
0, & \text{otherwise}
\end{cases}$$

In the above maximization, arbitrary tie-breaking can be performed if necessary. Next, inserting $A'_{mn}(\lambda)$ into (4.23), we have the linear optimization problem

$$\max_{X} \sum_{m=1}^{N} \sum_{n=1}^{N} x_{mn} A'_{mn}(\lambda)$$

subject to (4.24).

It is well known that there always exists an optimal solution to (4.28) that is binary [20, Chapter 3]. Then, to find a binary optimal $X$, (4.28) is a two-dimensional assignment problem.
Efficient algorithms, such as the Hungarian Algorithm [35], exist to produce an optimal solution with computational complexity being polynomial in $N$.

Finally, the optimal $\tilde{\phi}_{mnk}$ given $\lambda$ is

$$\tilde{\phi}^*_{mnk}(\lambda) = x^*_{mn}(\lambda)y^*_{mn}(\lambda).$$ \hspace{1cm} (4.29)

Since binary $x^*_{mn}(\lambda)$ and $y^*_{mn}(\lambda)$ are computed following the above procedure, $\tilde{\phi}^*_{mnk}(\lambda)$ is also binary. This shows that there exists at least one binary optimal solution to the maximization in (4.21). We offer several remarks on this solution in the following:

- Intuitively, the globally optimal solution described above suggests a pairing between the input and output channels at the relay, and if channels $m$ and $n$ are paired, they are assigned to a single user $k$, whose associated $A_{mnk}(\lambda)$ is the greatest among all users. Note that such an interpretation might lead us to conclude that we could have forgone continuous relaxation from the very beginning and focused only on a binary $\Phi$. However, we would still have required the continuous $\tilde{\Phi}$ to construct a convex optimization problem, whose strong duality property provides the optimality of the proposed approach. The optimality of $(X, \{y^{mn}\})$ taking binary values is implied only through the above derivation.

- Interestingly, the original optimization problem (4.21) with a binary matrix $\Phi$ is a special case of the axillary three-dimensional assignment problems. As implied in Section 2.2.2, the general form of this family of problems is NP hard and cannot be solved by continuous relaxation on $\Phi$, unlike the two-dimensional assignment problem in (4.28). In our case, the special structure of $\tilde{\Phi}$ expressed in (4.11), namely the absence of a constraint on per-user resource allocation, makes possible the availability of an efficient solution to (4.21).

- It is also worth noting that, given any $\lambda$, there may exist non-integer optimal solutions to (4.21). For example, when the same maximal value of $A_{mnk}$ is achieved by multiple
users in (4.26), there is an infinite number of optimal $y_k^{mn}$ values that maximize $A'_mn$, leading to non-integer optimal solutions for $\tilde{\phi}_{mnk}$. However, the procedure above finds only one of the optimal solutions in binary form, which is sufficient.

### 4.2.4 Dual Minimization: Baseline Subgradient Approach

The previous subsection provides a way to find the Lagrange dual $g(\lambda)$ for any Lagrange multiplier vector $\lambda$. Next, the standard approach calls for minimizing the dual function (refer to Section 2.2):

$$\min_{\lambda} g(\lambda)$$

s.t. $\lambda \succeq 0$.

This can be solved using the subgradient method described in Section 2.2.3. It is easy to verify that a subgradient at the point $\lambda$ is given by

$$\theta(\lambda) = \left( P_s - \sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{k=1}^{K} P_{mnk}^{ss}(\lambda), P_r - \sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{k=1}^{K} P_{mnk}^{rs}(\lambda), P_t - \sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{k=1}^{K} (P_{mnk}^{ss}(\lambda) + P_{mnk}^{rs}(\lambda)) \right),$$

(4.31)

where $P_{mnk}^{ss}(\lambda)$ and $P_{mnk}^{rs}(\lambda)$ are computed based on (4.18) and $\tilde{\phi}_{mnk}^*(\lambda)$ found using (4.29).

For completeness, we first summarize the standard subgradient updating algorithm for solving the dual problem in the following. We will present a modified dual minimization algorithm in Section 4.2.6, which is guaranteed to converge in polynomial time.

1. Initialize $\lambda^{(0)}$.

2. Given $\lambda^{(l)}$, obtain the optimal values of $P_{mnk}^{ss}(\lambda^{(l)})$, $P_{mnk}^{rs}(\lambda^{(l)})$, and $\tilde{\phi}_{mnk}^*(\lambda^{(l)})$.

3. Update $\lambda$ through $\lambda^{(l+1)} = [\lambda^{(l)} - \theta(\lambda^{(l)})\nu^{(l)}]^+$ where $\nu^{(l)}$ is the step size at the $l$th iteration.

4. Let $l = l + 1$; repeat from Step 2) until the convergence of $\min_l g(\lambda^{(l)})$. 


Referring to Section 2.2.3 regarding the step size topic, here we make use of constant step size $\nu$, i.e., $\nu^{(l)} = \nu$, or a constant step length $\nu$, i.e., $\nu^{(l)} = \nu / \| \theta(\lambda^{(l)}) \|_2$. They lead to an objective within a given neighborhood of a global optimum; while using the non-summable, square-summable rule leads to asymptotic convergence to a global optimum. Furthermore, one may satisfy any constraints on $\lambda^{(l)}$ within a convex region by projecting $\lambda^{(l)}$ onto the region. This is the general projected subgradient method, which does not reduce the speed of convergence [38]. For example, Step 3 above ensures that $\lambda^{(l)} \succeq 0$, and we will further consider projection onto convex regions $R_1$ and $R_2$ in Section 4.2.6.

### 4.2.5 Primal Optimality

With standard subgradient updating, the dual optimal $\lambda^*$ is obtained, from which we compute the channel assignment and power allocation matrices $(\Phi^*, P_s^*, P_r^*)$, where $\Phi^* = \tilde{\Phi}^*$. Since the optimization problem (4.10) is a convex program that satisfies Slater’s condition, it has zero duality gap. Denote by $f^*(\pi)$ the maximal value of the objective in (4.10). Then $f^*(\pi) = g(\lambda^*)$, and furthermore it is concave in $\pi$. We consider systems that have the following strictly diminishing rate-power relation:

**Assumption 1.** $f^*(\pi)$ is strictly concave in any strictly active power constraint $P_x \in \{ P_s, P_r, P_t \}$.

In other words, as the data rate increases, each unit of increment requires more and more marginal power. With a strictly concave $R(m, n, k)$ in terms of $p_{mnk}$, this assumption holds when either there is no tie-breaking in (4.27) or (4.28) or there is tie-breaking that is due to users or paths having the same weights or channel gains.$^2$

**Proposition 4.1.** Under Assumption 1, $(\Phi^*, P_s^*, P_r^*)$ is a globally optimal solution to the original problem (4.8).

---

$^2$However, we cannot rule out the possibility of a case where other forms of tie-breaking in (4.27) or (4.28) might create linear segments in $f^*(\pi)$, although in all simulation tests with arbitrary parameters, we have not produced a case where this assumption fails.
Proof. Since $P^{ss}$ and $P^{r*}$ are uniquely determined by $\lambda^*$ and $\Phi^*$, we need only to focus on $\Phi^*$.

For any inactive constraint $P_x$, we have $\lambda^*_x = 0$ and the subgradients of $g(\lambda)$ in the direction of $\lambda_x$ are all positive. Hence any $\Phi^*$ is feasible with respect to $P_x$.

For any strictly active constraint $P_x$, we have $\lambda^*_x > 0$. Furthermore,

$$f^*(\pi) = \mathcal{L}(\Phi^*, P^{ss}, P^{r*}, \lambda^*)$$

$$\leq f^*(\sum_{m,n,k} p^{*}_{mnk}) - \lambda^*(\sum_{m,n,k} p^{*}_{mnk} - \pi)^T \tag{4.32}$$

Given Assumption 1, the above is possible only when all strictly active constraints are satisfied with equality. Therefore, any $\Phi^*$ is feasible with respect to $P_x$ and the complementary slackness condition is satisfied. Hence, $(\Phi^*, P^{ss}, P^{r*})$ is a globally optimal solution to (4.10).

Furthermore, since (4.10) is a constraint-relaxed version of (4.8), $\Phi^*$ gives an upper bound to the objective of (4.8). Finally, since $\Phi^*$ satisfies the binary constraints in (4.8) at each iteration of the subgradient algorithm, it satisfies all constraints in (4.8). Therefore, it is a globally optimal solution to (4.8).

We point out that using conventional convex optimization software packages directly on the relaxed problem (4.10) is not sufficient to solve (4.8). This is because there is no guarantee that they will return a binary $\tilde{\Phi}^*$, and furthermore due to complicated three-dimensional dependencies among $\phi_{mnk}$, there is no readily available method to transform a fractional $\tilde{\Phi}^*$ to the desired binary solution.

### 4.2.6 Dual Minimization: Divide-and-Conquer Algorithm

The standard subgradient method produces a global optimum, but its computational complexity is not generally known. Previous studies have provided asymptotic bounds or conjectures on its efficiency through computational experience. In general, the number of iterations in subgradient updating depends on the step-size rule, the distance between the initial solution and the optimal solution, and the 2-norm of the subgradients [37][38].
Next, we propose a new dual minimization algorithm that guarantees convergence with polynomial complexity in \( N \) and \( K \), to a global optimum for our optimization problem. It uses a divide-and-conquer approach, by grouping the possible locations of \( \lambda^* \) into two regions and applying projected subgradient updating constrained within either. It ensures that in each region, our choice of the initial \( \lambda^{(0)} \) and subsequent subgradient updating lead to convergence in polynomial time.

We first define the following two overlapping convex regions in terms of \( \lambda \):

\[
\mathcal{R}_1 \triangleq \left\{ \lambda : \lambda_s + \lambda_t \geq \frac{\min_{\{k:w_k>0\}} w_k \min \left\{ \min_{\{m:a_m>0\}} a_m, \min_{\{m,k:c_{mk}>0\}} c_{mk} \right\}}{4\alpha \left( \max_m a_m \min \{P_s, P_t\} + 1 \right)}, \lambda \succeq 0 \right\},
\]

\[
\mathcal{R}_2 \triangleq \left\{ \lambda : \lambda_s + \lambda_t + \frac{\min_{\{m:a_m>0\}} a_m}{\max_{n,k} b_{nk}} (\lambda_r + \lambda_t) \geq \frac{\min_{\{k:w_k>0\}} w_k}{\alpha \left( \min \{P_s, P_t\} + \frac{1}{\min_{\{m:a_m>0\}} a_m} \right)}, \lambda \succeq 0 \right\}.
\]

These are two possible regions where \( \lambda^* \) resides, which depends on whether there exists at least one *chosen* path with non-zero direct-link channel gain \( c_{mk} \). This is formalized in the following lemma.

**Lemma 4.2.** If there exists some \((m, n, k)\) such that \( \phi_{mnk}^* = 1 \) and \( c_{mk} > 0 \), then \( \lambda^* \in \mathcal{R}_1 \). Otherwise, \( \lambda^* \in \mathcal{R}_2 \).

*Proof.* The proof is provided in Appendix B 7.2.2. \( \square \)

The proposed divide-and-conquer dual minimization (DCDM) algorithm considers both possible regions for \( \lambda^* \). It first creates the two conditions in Lemma 4.2 by artificially setting direct-link channel gains to zero. It then applies the projected subgradient algorithm on \( \mathcal{R}_1 \) and \( \mathcal{R}_2 \) separately, and chooses the better solution between these two. The algorithm is formally detailed in Algorithm 1, and its optimality and complexity are given in Propositions 4.2 and 4.3, respectively. Note that one cannot use Lemma 4.2 to determine, *before* the optimal channel assignment matrix \( \tilde{\Phi} \) is chosen, which region \( \lambda^* \) is in. This necessitates the comparison step in the DCDM algorithm.
Algorithm 1 Divide-and-Conquer Dual Minimization (DCDM)

\textbf{if} there exists some \( m \) and \( k \) such that \( c_{mk} > 0 \) \textbf{then}

\[ \lambda_1^* = \text{output of subgradient updating algorithm with projection onto } \lambda^{(l)} \in \mathcal{R}_1 \]

Set \( c_{mk} = 0 \) for all \( 1 \leq m \leq N \) and \( 1 \leq k \leq K \)

\[ \lambda_2^* = \text{output of subgradient updating algorithm with projection onto } \lambda^{(l)} \in \mathcal{R}_2 \]

\textbf{return} \( \arg\min_{\lambda \in \{\lambda_1^*, \lambda_2^*\}} g(\lambda) \)

\textbf{else}

\[ \lambda^* = \text{output of subgradient updating algorithm with projection onto } \lambda^{(l)} \in \mathcal{R}_2 \]

\textbf{return} \( \lambda^* \)

\textbf{end if}

Proposition 4.2. With DCDM, the computed channel assignment and PA matrices \((\Phi^*, P^*, P^{**})\), where \( \Phi^* = \tilde{\Phi}^* \), is a globally optimal solution to the original problem (4.8).

\textbf{Proof.} Suppose there exists some \((m, n, k)\) such that \( \phi_{mnk}^* = 1 \) and \( c_{mk} > 0 \). Then Lemma 4.2 shows that \( \lambda^* \in \mathcal{R}_1 \). Therefore, by Proposition 4.1, \( \lambda_1^* \) obtained by subgradient updating projected onto \( \mathcal{R}_1 \) is an optimal solution. Furthermore, setting \( c_{mk} = 0 \) for all \( 1 \leq m \leq N \) and \( 1 \leq k \leq N \) only reduces \( R(m, n, k) \) for all paths, so that subsequently minimizing the Lagrange dual yields an inferior solution. Therefore, \( \arg\min_{\lambda \in \{\lambda_1^*, \lambda_2^*\}} g(\lambda) = \lambda_1^* \) is returned by DCDM.

Suppose \( c_{mk} = 0 \) for all \((m, n, k)\) such that \( \phi_{mnk}^* = 1 \), i.e., all chosen paths have zero direct-link channel gain. Then, setting \( c_{mk} = 0 \) for all \( 1 \leq m \leq N \) and \( 1 \leq k \leq N \) only reduces \( R(m, n, k) \) for the non-chosen paths. Subsequently minimizing the Lagrange dual yields the same solution as before changing \( c_{mk} \). Furthermore, Lemma 4.2 shows that this optimal solution is in \( \mathcal{R}_2 \). Hence, \( \lambda_2^* \) obtained by subgradient updating projected onto \( \mathcal{R}_2 \) is an optimal solution. In this case, \( \arg\min_{\lambda \in \{\lambda_1^*, \lambda_2^*\}} g(\lambda) = \lambda_2^* \) is returned by DCDM.

The polynomial computational complexity of DCDM is stated in Proposition 4.3. Its proof requires the following lemmas, which give upper bounds on \( \|\lambda^*\|_2 \) and \( \|\theta(\lambda^{(l)})\|_2 \), where \( \|\cdot\|_2 \)
denotes the 2-norm.

**Lemma 4.3.** At global optimum, $\|\lambda^*\|_2$ is upper bounded by $\lambda_{\text{max}} = O(N^2)$.

*Proof.* The proof is provided in Appendix B 7.2.3. □

**Lemma 4.4.** At every step of subgradient updating in the DCDM algorithm, $\|\theta(\lambda(l))\|_2$ is upper bounded by $\theta_{\text{max}} = O(N^2)$.

*Proof.* The proof is provided in Appendix B 7.2.4. □

**Proposition 4.3.** To achieve a weighted sum-rate within an arbitrary $\epsilon > 0$ neighborhood of the optimum $g(\lambda^*)$, using either a constant step size or a constant step length in subgradient updating, the DCDM algorithm has polynomial computational complexity in $N$ and $K$.

*Proof.* At each iteration of the standard subgradient updating algorithm, the procedures described in Sections 4.2.2 and 4.2.3 are employed. This has computational complexity polynomial in $N$ and $K$. Therefore, it remains to show that the total number of iterations is not more than polynomial in $N$ or $K$.

For either case of projecting onto $R_1$ or $R_2$, one may choose an initial $\lambda^{(0)}$ such that the distance between $\lambda^{(0)}$ and $\lambda^*$ is upper bounded by $\lambda_{\text{max}}$. Then, it can be shown that, at the $l$th iteration, the distance between the current best objective to the optimum objective $g(\lambda^*)$ is upper bounded by $\lambda_{\text{max}}^2 + \nu^2\theta_{\text{max}}^2/l$ if a constant step size is used (i.e., $\nu(l) = \nu$), or by $\lambda_{\text{max}}^2 + \nu^2\theta_{\text{max}}^2/l$ if a constant step length is used (i.e., $\nu(l) = \nu/\|\theta(\lambda(l))\|_2$) [37, 38]. For the former and latter bounds, if we set $\nu = \epsilon/\theta_{\text{max}}^2$ and $\nu = \epsilon/\theta_{\text{max}}$ respectively, both are upper bounded by $\epsilon$ when $l \geq \lambda_{\text{max}}^2\theta_{\text{max}}^2/\epsilon^2 = O(N^4)$. Hence, the number of required iterations until convergence, for either of the two projected subgradient updating procedures in DCDM, is polynomial in $N$ and independent of $K$. □

Note that, in practice, using the non-summable, square-summable step-size rule, e.g., $\nu(l) = \frac{c}{l}$ for some positive constant $c$, in the early iterations of subgradient updating, often leads to faster movement toward a global optimum than using a constant step size or a constant step size.
length. This is due to its larger step sizes when \( l \) is small. However, such a step-size rule does not guarantee polynomial convergence time\(^3\). Therefore, one may start with the non-summable, square-summable rule, and then switch to one of the constant-step rules when the step size or step length is sufficiently near the prescribed value in Proposition 4.3. This would reduce the convergence time in practice while preserving the guarantee of polynomial complexity.

### 4.3 Extensions to General Relaying Strategies

For any relaying strategy in which data sent through different communication paths \( P(m, n, k) \) are independent and the achievable rates \( R(m, n, k) \) is a concave function in transmission powers \((P_{mnk}^s, P_{mnk}^r)\), the proposed solution approach gives jointly optimal channel assignment and PA for weighted sum-rate maximization. To see this, we first note that any concave rate function would lead to convex programming for the relaxed and reformulated problem, which satisfies Slater’s condition and hence has zero duality gap. Furthermore, toward maximizing the Lagrange function, we can generalize (4.16) into the following form:

\[
\max_{P_{mnk}^s \geq 0, P_{mnk}^r \geq 0} w_k \tilde{\phi}_{mnk} R\left( \frac{P_{mnk}^s}{\tilde{\phi}_{mnk}}, \frac{P_{mnk}^r}{\tilde{\phi}_{mnk}} \right) - (\lambda_s + \lambda_t) P_{mnk}^s - (\lambda_r + \lambda_t) P_{mnk}^r.
\] (4.33)

Since the partial derivatives of the above maximization objective contains \( P_{mnk}^s \) and \( P_{mnk}^r \) only in the form of \( \frac{P_{mnk}^s}{\tilde{\phi}_{mnk}} \) and \( \frac{P_{mnk}^r}{\tilde{\phi}_{mnk}} \), we always have \( P_{mnk}^{ss} \) and \( P_{mnk}^{rr} \) as the product of \( \tilde{\phi}_{mnk} \) and a non-negative factor. This leads to a maximization problem of the form in (4.21), which has been shown to admit a binary optimal solution in Section 4.2.3.

Besides DF, the time-sharing variants of any relaying strategies with long-term or short-term average power constraints, as well as all capacity achieving strategies, have concave achievable rates [86]. Our algorithm is applicable to these current and future relaying strategies to find the optimal solution. However, the closed-form solutions for \((P_{mnk}^{ss}, P_{mnk}^{rr})\) may be difficult to find in some cases, requiring more involved numerical computation.

---

\(^3\)Consider the following idealized example for illustration. If \( \nu(l) = \frac{1}{l} \) for all \( l \), the number of iterations would need to be \( L = \Theta(e^{\lambda_{max}}) \) to satisfy the convergence requirement \( \sum_{l=1}^{L} \nu(l) = \Theta(\lambda_{max}) \).
For relaying strategies that do not have concave achievable rates, such as AF, near-optimal solutions can be obtained by using the proposed approach in the following senses:

1. A concave bound of the achievable rate may be used to approximate $R(m, n, k)$. For example, with AF, we have
   \[ R(m, n, k) = \frac{1}{2} \log \left( 1 + \frac{a_m b_n P_r^m P_s^m P_s^m}{1 + \frac{a_m b_n P_r^m P_s^m}{1 + b_n P_r^m P_s^m}} + P_s^m c_{mk} \right). \]
   A concave upperbound is obtained by removing “1” from the denominator. By substituting such a concave bound for $R(m, n, k)$ in the original optimization problem, we obtain a solution that optimizes in terms of the bound. In the case of AF, such solution is near-optimal for weighted sum-rate, since the “1” is negligible for paths with high effective SNR, while paths with low effective SNR do not contribute substantially to the performance objective.

2. It has been shown in [34] that, regardless of the convexity of the objective function in a multi-channel resource assignment problem, if the objective at optimum is a concave function of the maximum allowed powers, the duality gap of the Lagrange dual induced by power constraints is zero. This is due to time-sharing over resource assignment strategies. Furthermore, there is a frequency-domain approximation of time-sharing, so that the duality gap is asymptotically zero when the number of channels goes to infinity. Hence, for systems with a large number of channels, near-optimal results can be achieved by the proposed approach.

### 4.4 Numerical Results

In this section, we compare the performance of jointly optimal channel pairing, channel-user assignment, and PA with that of suboptimal schemes. We further study the different factors that affect the performance gap under these schemes, in order to shed light on the tradeoff between performance optimality and implementation complexity. The suboptimal schemes considered are
- **No Pairing**: Channel-user assignment and PA are jointly optimized, but no channel pairing is performed, i.e. the same incoming and outgoing channels are assumed. The solution is found by always assigning an identity matrix to $X$ instead of solving (4.28).

- **No PA**: Allocate power uniformly across all channels, subject to power constraints. Channel pairing and channel-user assignment are jointly optimized by solving (4.8) with given power, which is a three-dimensional assignment problem over binary $\Phi$. The solution is found by following the procedure in Section 4.2.3.

- **Separate Optimization**: The three-stage solution proposed in [69], with channel-user assignment based on maximum channel gain over the second-hop channels, channel pairing based on sorted channel gains, and water-filling PA.

- **Max Channel Gain**: Channel-user assignment by maximum channel gain over the second hop, with uniform PA and no channel pairing.

We use OFDMA as an example for a multi-channel system. The relaying network setup is shown in Fig. 4.2, where the distance between the source and the relay is denoted by $d_{sr}$, and $K = 4$ users are located on a half-circle arc around the relay with radius $d_{rd}$. A 4-tap frequency-selective propagation channel is assumed for each hop, and the number of channels is set to $N = 16$. We define a nominal SNR, denoted by $\text{SNR}_{\text{nom}}$, as the average received SNR over each subcarrier under uniform PA. Specifically, with total power constraint $P_t$, we have $\text{SNR}_{\text{nom}} \triangleq \frac{P_t (\bar{d}_{sd})^{-\kappa}}{2 \sigma^2 N}$, where $\kappa = 3$ denotes the pathloss exponent, $\sigma^2$ denotes the noise power per channel, and $\bar{d}_{sd}$ denotes the average distance between the source and users. A total power constraint and equal individual power constraints on both the source and the relay are assumed with $P_s = P_r = \frac{2}{3} P_t$, unless it is stated otherwise.

**Performance versus Nominal SNR**

We compare the performance of various channel assignment and PA schemes at different $\text{SNR}_{\text{nom}}$ levels for $K = 4$. We fix the ratio $d_{sr}/d_{rd}$ to be $1/3$. Fig. 4.3 depicts the normalized weighted sum-rate (normalized over $N$) vs. $\text{SNR}_{\text{nom}}$ for DF relaying with equal weight, i.e.,
CHAPTER 4. CHANNEL AND POWER ASSIGNMENT FOR MULTI-USER RELAYING

Figure 4.2: Simulation configuration with \( K = 4 \) users

![Simulation configuration with 4 users](image)

Figure 4.3: Normalized weighted sum-rate vs. nominal SNR with \( w = [0.25, 0.25, 0.25, 0.25] \), \( N = 16, K = 4 \), and DF relaying.

\[
\mathbf{w} \triangleq [w_k]_{1 \times K} = [0.25, 0.25, 0.25, 0.25].
\]

The jointly optimal scheme outperforms the other sub-optimal schemes, and provides as much as 20\% gain over the Separate Optimization scheme. The gain is increased when an unequal weight vector \( w \) is required to satisfy different user QoS demands or fairness. Fig. 4.4 shows the normalized weighted sum-rate vs. SNR\(_{\text{nom}}\) for \( w = [0.15, 0.15, 0.35, 0.35] \), where a substantial gain is observed by employing the jointly optimal solution.

**Performance versus Number of Users**

In this experiment, we show how the number of users affects the performance of various resource assignment schemes. We increase the number of users in Fig 4.2, and uniformly place
CHAPTER 4. CHANNEL AND POWER ASSIGNMENT FOR MULTI-USER RELAYING

Figure 4.4: Normalized weighted sum-rate vs. nominal SNR with $w = [0.15, 0.15, 0.35, 0.35]$, $N = 16$, $K = 4$, and DF relaying.

them around the half-circle arc. In order to properly compare weighted sum-rate under different number of users, we do not normalize $\sum_{k=1}^{K} w_k = 1$. Instead, we fix $w_k = 1$ for all $k$. The nominal SNR is $\text{SNR}_{\text{nom}} = 4$ dB, and the ratio $d_{sr}/d_{rd} = 1/3$. Fig 4.5 shows the normalized weighted-sum rate vs. the number of users for DF relaying under total power constraint $P_t$. As we see, the sum-rate is improved due to the multi-user diversity gain with an increased number of users. In addition, consistent performance gain under joint optimization can be seen over different user population sizes.

**Impact of Relay Position**

Through this experiment, we study how the relay position affects the performance under various resource assignment schemes. The $K = 4$ users are located close to each other as a cluster, and they have approximately the same distance to the relay and the source. We change the relay position along the path between the source and the user cluster. Figs 4.7 and 4.8 demonstrate the normalized weighted-sum rate vs. the ratio $d_{sr}/d_{sd}$. We set $w = [0.15, 0.15, 0.35, 0.35]$, and $\text{SNR}_{\text{nom}} = 3$ dB. Fig. 4.7 shows the DF relaying case under both total
We see from Fig. 4.7 that better performance is observed when the relay is closer to the source than to the users in DF relaying, as correctly decoding data at relay is important in successful DF relaying. In addition, comparing the joint optimal scheme with No Pairing scheme, we see that the gain of channel pairing is evident when the relay is closer to the source, but diminishes when the relay moves closer to the users. In the latter case, as the first-hop becomes the bottleneck, channel pairing at the second-hop provides no benefit. This is not the case for AF relaying. As shown in Fig. 4.8, channel pairing gain is observed throughout different relay positions. Furthermore, the performance of the jointly optimal solution only has mild variation throughout different relay positions, unlike the No PA scheme. This suggests that the benefit of optimal PA for AF relaying is more significant when the relay is closer to either the source or the users. Note that it is not surprising that No Pairing scheme outperforms Separate Opt. approach because the former scheme adaptively allocates the second-hop channels to the
users taking into account the users’ weights, while the latter scheme uses max-SNR channel assignment which is independent of the users’ weight.

**Search Complexity vs Number of Channels** Fig. 4.6 gives an example that illustrates the complexity vs. the number of channels. A multi-channel downlink system with $K = 4$ users with non-equal user weights is considered. For different $N$ under the same total power constraint, we plot the number of update iterations required for the subgradient method with constant step size $\nu = 5 \times 10^{-4}$ and initial point $\lambda^0 = 0.1$ to reach to $\epsilon = 0.5\%$ neighborhood of the total available power i.e., $|\sum_m \sum_n \sum_k \phi^*_{mnk}(\lambda^l) (P_{ss}^{mnk}(\lambda^l) + P_{rr}^{mnk}(\lambda^l)) - P_t|/P_t \leq 5 \times 10^{-3}$. Note that instead of dual function $g(\lambda^*)$ for termination condition, we equivalently used the total power function by invoking the fact that at optimality the power constraint is satisfied with equality i.e., $\sum_m \sum_n \sum_k \phi^*_{mnk} (P_{ss}^{mnk} + P_{rr}^{mnk}) = P_t$. As anticipated, this figure clearly shows that the overall complexity is polynomial.

![Figure 4.6: Number of iterations vs. number of channels for DF OFDMA relaying with $K = 4$, equal non-equal weight under total power constraint](image-url)
Figure 4.7: Normalized weighted sum-rate vs. relay location; $K = 4$, $w = [.15, .15, .35, .35]$, $N = 16$, and DF relaying.

Figure 4.8: Normalized weighted sum-rate vs. relay location; $K = 4$, $w = [.15, .15, .35, .35]$, $N = 16$, and AF relaying.
4.5 Summary

In this chapter, we studied the problem of jointly optimizing CP, channel-user assignment, and PA in a general single-relay multi-channel multi-user system. Although such joint optimization naturally leads to a mixed-integer programming problem which typically has prohibitive complexity to solve, it is shown that there is an efficient algorithm to find an optimal solution in our problem. The proposed approach is built on convex relaxation and transforms the original problem into a specially structured three-dimensional assignment problem, which not only preserve the binary constraints and strong Lagrangian duality of the continuous version, but in some cases can also lead to polynomial-time computation complexity through careful choices of the optimization trajectory. It is argued that the proposed framework is applicable to a wide variety of scenarios, accommodating different relaying strategies, total and individual power constraints, and maximum weighted sum-rate or max-min rate objectives.
Chapter 5

Jointly Optimal Bit Loading, CP and PA for Multi-channel Relaying

In this chapter, we design a jointly optimal channel pairing, power allocation and bit loading scheme for multi-channel relaying, to maximize the overall bit rate under a single power constraint. The power constraint can be on the source, the relay, or both as an upper limit to their total power. As we will see, the resulting optimization problem is of non-convex mixed-integer programming in nature, with both a discrete objective and a discrete feasible region. Although the Lagrange duality approach can be applied to such problems [22], it generally does not lead to an optimal solution due to the non-zero duality gap. It is shown in [34] that, for many non-convex problems in broadband multi-channel resource allocation, a time-sharing property – a concavity condition of the objective function with respect to the total power budget – still ensures zero-duality gap, and thus the problem can be solved by the Lagrange dual. This result has been found useful to provide asymptotically optimal solution to some resource allocation problem in OFDM systems [56][64]. In our problem, however, the discrete objective function renders such condition invalid, and we show that a non-zero duality gap exists.

Nonetheless, we develop our solution through the Lagrange dual approach. By exploring the structure of our problem, we are able to bound the gap to the original objective to be within...
CHAPTER 5. JOINTLY OPTIMAL BIT LOADING, CP AND PA FOR MULTI-CHANNEL RELAYING

one bit. This knowledge allows us to extract the exact optimal integer solution. We further develop numerical techniques to reduce the algorithm runtime.

The rest of this chapter is organized as follows. In Section 5.1, we describe the system model and formulate the joint optimization problem. In Section 5.2, we describe the Lagrange dual approach that produces an upper bound to our optimal objective, and in Section 5.3, we present a study on the special structure of our problem that enables the extraction of the required integer optimal solution. We further present a complexity reduction method and show that the proposed algorithm has polynomial runtime in Section 5.4. Numerical results are given in Section 5.5, followed by a summary in Section 5.6.

5.1 System Model and Problem Statement

We consider a three-node cooperative scenario where a source communicates with a destination with the assistance of a relay. Adopting the same notations defined in Section 2.1.4 let \( h_{1n} \) and \( h_{2n} \) for \( n = 1, \ldots, N \) denote the channel gain on channel \( n \) over the first hop and the second hop, respectively. In the system formulation, we always use the normalized channel power gain, which for channel \( n \) of the first and second hops are respectively defined as \( a_n \triangleq \frac{|h_{1n}|^2}{\sigma_r^2} \) and \( b_n \triangleq \frac{|h_{2n}|^2}{\sigma_d^2} \), where \( \sigma_r^2 \) and \( \sigma_d^2 \) are the noise variances at the relay and the destination, respectively.

We assume two-phase cooperative transmission as described in the background chapter. We are not limited to any specific relaying strategy (DF, AF, etc), as long as the rate-power relation is known and satisfies some conditions as explained below. Nonetheless, the relay is supposed to conduct channel pairing, matching each incoming channel with an outgoing channel. Recalling the binary pairing indicator \( \phi_{mn} \) and channel pairing matrix \( \Phi = [\phi_{mn}]_{N \times N} \) from section 2.1.3, we have the following constraints:

\[
\sum_{n=1}^{N} \phi_{mn} = 1, \quad \forall m; \quad \sum_{m=1}^{N} \phi_{mn} = 1, \quad \forall n.
\]  

(5.1)

We assume there is a single power constraint \( P_t \). It can be the total power constraint on
the source and relay transmissions, which may be due to, e.g., hardware limitations or legal mandates on the overall interference created by the system. It can also be an individual power constraint imposed on the source or the relay.

Let \( p_{mn} \) be the part of \( P_t \) adaptively allocated to path \((m, n)\), and define the power allocation matrix \( P = [p_{mn}]_{N \times N} \). The power constraint is then expressed by

\[
0 \leq p_{mn} \leq P_t,
\]

\[
\sum_{m=1}^{N} \sum_{n=1}^{N} \phi_{mn} p_{mn} \leq P_t.
\]

**Bit Loading**

For each path, the data transmission may adopt varying rates based on the modulation or coding scheme in use. We define a bit loading matrix \( B = [b_{mn}]_{N \times N} \), where \( b_{mn} \) denotes the number of bits per transmission over path \((m, n)\). We assume \( b_{mn} \) takes only non-negative integral values, i.e.,

\[
b_{mn} \in \mathcal{M},
\]

where \( \mathcal{M} = \{0, 1, 2, \ldots, M\} \). For example, in \( x\)-QAM, we have \( x = 2^{b_{mn}} \) for \( b_{mn} = 1, 2, \ldots \).

Note that \( b_{mn} = 0 \) corresponds to the case where either path \((m, n)\) is not selected or not allocated sufficient power for useful transmission.

In order to support the transmission of \( b_{mn} \) bits, the allocated power \( p_{mn} \) must satisfy

\[
b_{mn} \leq R_{mn}(p_{mn}),
\]

where \( R_{mn} \) is in general a real-valued function that depends on the channel power gains \( a_m \) and \( b_n \). Moreover, other system parameters as well as performance requirements may be involved in constructing \( R_{mn} \), such as the modulation and coding schemes, the relaying strategy, and the required error rate constraints.

For instance, with a repetition-coding based DF relaying strategy, given a total power constraint \( p_{mn} \) between the source and the relay on path \((m, n)\), we may have \( R_{mn}(p_{mn}) = \frac{1}{2} \log(1 + \Gamma^{-1} a_{mn} p_{mn}) \), where \( a_{mn} = (\frac{1}{a_m} + \frac{1}{b_n})^{-1} \), and \( \Gamma \) is an SNR gap corresponding to the
distance between the theoretic capacity and the achievable rate due to modulation and coding [87]. As another example, for AF relaying under relay power constraint \( p_{mn} \), the achievable rate is given by

\[
R_{mn}(p_{mn}) = \frac{1}{2} \log(1 + \Gamma - \frac{1}{p_s m a m p_{mn} b_n (1 + p_s m a m + p_{mn} b_n)})
\]

where \( p_s \) denotes the given source power on channel \( m \) of the first hop.

In this study, we consider relaying strategies whose \( R_{mn}(p_{mn}) \) is a strictly increasing and concave function of \( p_{mn} \). Both examples above can be shown to satisfy this assumption.

### 5.1.1 Optimization Objective

Our objective is to design a scheme that maximizes the overall source-destination bit rate. It is cast in the following form:

\[
B^*(P_t) = \max_{\Phi, b, p} \sum_{m=1}^{N} \sum_{n=1}^{N} \phi_{mn} b_{mn}
\]

\[
\text{s.t. (2.4), (5.1), (5.2), (5.3), (5.4), (5.5)}
\]

For given total power budget \( P_t \), the optimization problem (5.6) seeks the jointly optimal solution for channel pairing, power allocation, and bit loading by optimizing \( (\Phi, b, p) \). In our model independent encoding/decoding scheme is assumed across the subchannels. Moreover, worthwhile to mention that our objective here is different from the information-theoretic rate that we have considered so far in the previous chapters of this thesis. In this chapter, we concern the data rate or throughput of the system that always adopts integer value for the given level of error probability.

For convenience of exposition in the rest of this chapter, we also consider the minimum required power given an integer sum rate target \( B \):

\[
P_t^*(B) = \min_{\Phi, b, p} \sum_{m=1}^{N} \sum_{n=1}^{N} \phi_{mn} p_{mn}
\]

\[
\text{s.t. (2.4), (5.1), (5.2), (5.4), (5.5)}
\]

\[
\sum_{m=1}^{N} \sum_{n=1}^{N} \phi_{mn} b_{mn} = B
\]
We note that $P^*_t(k) < P^*_t(l)$ for all $k < l$, so $B^*(P_t)$ is an integer-valued increasing staircase function in $P_t$. Let $\Delta P^*_t(k) \triangleq P^*_t(k+1) - P^*_t(k)$. We consider systems that have the following strictly diminishing rate-power relation:

$$\Delta P^*_t(k) > \Delta P^*_t(k-1), \quad \forall k \geq 0. \quad (5.8)$$

This reflects the general reality in practical systems that, as the data rate increases, each one-bit increment requires more and more marginal power. This can also be viewed as a discrete and strict version of the time-sharing property, which is widely applicable in broadband wireless resource management [34].

The optimization problem (5.6) is non-convex, with integer constraints on both the objective function and the feasibility region. Such problems usually is NP-hard owing to the combinatorial nature of channel pairing and bit loading. The Lagrange dual decomposition method is a commonly used approach to tackle this type of problems. In a recent study, the authors have found an efficient method to jointly optimize channel assignment and power allocation for systems without bit loading [18], showing that a zero Lagrangian duality gap can be achieved even with discrete channel pairing. However, the discrete nature of the objective function in (5.6) renders that method inapplicable.

Nonetheless, the Lagrange dual decomposition method provides an upper bound to the optimal solution. In Sections 5.2 and 5.3, we show that an optimal solution can be extracted from the minimizer of the Lagrange dual function, even though the problem under consideration has a positive duality gap. Furthermore, in Section 5.4 we observe that the proposed solution can be achieved within polynomial time in $N$, providing substantial savings in computing resources over brute-force search methods.
5.2 Upper Bound to Optimal Bit Rate via Dual Decomposition

Let $\lambda$ be the Lagrange multiplier associated with the power constraint (5.3). We have the following Lagrange function for (5.6):

$$L(\Phi, p, b, \lambda) = \sum_{m,n} \phi_{mn} b_{mn} - \lambda \left( \sum_{m,n} \phi_{mn} p_{mn} - P_t \right).$$  \hspace{1cm} (5.9)

The dual function $g(\lambda)$ is defined as the maximum of (5.9):

$$g(\lambda) = \max_{\Phi, p, b} L(\Phi, p, b, \lambda)$$ \hspace{1cm} (5.10)

subject to (2.4), (5.1), (5.2), (5.4), (5.5).

Then, due to the weak duality property of Lagrange decomposition [37], we can obtain an upper bound to $B^*(P_t)$ by taking the minimum of $g(\lambda)$, i.e.,

$$B^*(P_t) \leq B_{LD} \Delta = \min_{\lambda} \{ g(\lambda) : \lambda \geq 0 \}.$$ \hspace{1cm} (5.11)

The maximization and minimization problems of (5.10) and (5.11) can be solved in the following three consecutive steps.

5.2.1 Bit Loading, PA and CP Given Lagrange Multiplier, $\lambda$

Let $(\Phi^*(\lambda), b^*(\lambda), p^*(\lambda))$ be an optimal solution to (5.10). We first consider maximizing $L(\Phi, p, b, \lambda)$ over $b$ and $p$ for any given $\Phi$.

This problem can be decomposed into $N \times N$ underlying independent subproblems, concerning the optimal power and bit allocation for each possible path $(m, n)$.

$$\max_{p_{mn}, b_{mn}} b_{mn} - \lambda p_{mn}$$ \hspace{1cm} (5.12)

subject to

\begin{itemize}
  \item[i)] $b_{mn} \leq R_{mn}(p_{mn})$,
  \item[ii)] $p_{mn} \geq 0, \quad b_{mn} \in M$ .
\end{itemize}
Since $R_{mn}(p_{mn})$ is an increasing function of $p_{mn}$, it is not difficult to see that, at optimality, the constraint (i) in (5.12) must be satisfied by equality, i.e.,

$$p^*_mn(\lambda) = R^{-1}_{mn}(b^*_mn(\lambda)).$$

(5.13)

Furthermore, since $R_{mn}(p_{mn})$ is a strictly concave function, there is a unique maximizer $b^*_mn(\lambda)$ for (5.12). Substituting (5.13) into (5.12), we obtain a discrete concave function and can find $b^*_mn(\lambda)$ by using an efficient algorithm such as integer bisection search over the domain of $b_{mn}$.

Next, we substitute $b^*_mn(\lambda)$ and $p^*_mn(\lambda)$ into (5.10) and optimize it over $\Phi$, for the given $\lambda$. Let $L^*_mn(\lambda) = b^*_mn(\lambda) - \lambda p^*_mn(\lambda)$. Then the maximization in (5.10) is equivalent to

$$g(\lambda) = \max_{\Phi} \sum_{m,n} \phi_{mn} L^*_mn(\lambda)$$

(5.14)

s.t. (2.4), (5.1).

This optimization is a two-dimensional assignment problem [22], whose binary solution $\Phi^*(\lambda)$ can be efficiently obtained by several existing methods, such as the Hungarian Algorithm, which has computational complexity $O(N^3)$.

### 5.2.2 Dual Minimization via Subgradient Method

As explained in Section 2.2.3, the dual minimization problem (5.11) can be solved by employing the subgradient algorithm. For completeness, this baseline algorithm is summarized in the following.

1. Initialize $\lambda^{(0)}$.

2. Given $\lambda^{(l)}$, obtain $b^*_mn(\lambda^{(l)})$, $p^*_mn(\lambda^{(l)})$ and $\phi^*_mn(\lambda^{(l)})$ for all $m, n$ (Section 5.2.1).

3. Update $\lambda$ through $\lambda^{(l+1)} = \max\{\lambda^{(l)} - \theta(\lambda^{(l)})\nu^{(l)}, 0\}$ where we use the subgradient

$$\theta(\lambda) = P_t - \sum_{m,n} \phi^*_mn(\lambda) p^*_mn(\lambda),$$

(5.15)

and $\nu^{(l)}$ is the step size at the $l$th iteration.
4. Set \( l = l + 1 \); repeat from Step (2) until the convergence of \( \min_l g(\lambda^{(l)}) \).

We may adopt the square-summable, non-summable step-size rule [38]. Let \( \lambda^* \) be the global minimizer for the dual minimization problem (5.11).

The above Lagrange dual method is widely used to provide an upper bound to the optimal objective function. However, for general non-convex optimization problems, there is no guarantee that the primal solution corresponding to \( \lambda^* \) is feasible, so generally this method does not produce an optimal solution. As an alternative, in the general context of mixed-integer programming, the bound \( B_{LD} \) can be exploited by other algorithms such as branch-and-bound to find the optimal solution [22]. However, the application of such an approach usually requires exponential computational complexity due to the combinatoric nature of the problem.

Nonetheless, we next show a surprising observation that, in the problem under consideration, a jointly optimal solution for channel pairing, power allocation, and bit loading can be efficiently extracted, despite the non-zero duality gap.

### 5.3 Extraction of Jointly Optimal Solution

There is a fundamental difference between our proposed approach and the conventional Lagrange dual approaches that require a zero duality gap to derive an optimal solution [22]. In this section we show that a small perturbation of \( \lambda^* \) allows the extraction of the optimal solution to the primal problem (5.6).

In the subsequent derivations, we define

\[
B^{\text{tr}}(\lambda) \triangleq \sum_{m,n} \phi_{mn}^*(\lambda) b_{mn}^*(\lambda), \tag{5.16}
\]

\[
P^{\text{tr}}(\lambda) \triangleq \sum_{m,n} \phi_{mn}^*(\lambda) p_{mn}^*(\lambda). \tag{5.17}
\]

We further denote by \( \lambda^{*+} \) and \( \lambda^{*-} \), respectively, \( \lambda^* + d\lambda \) and \( \lambda^* - d\lambda \), where \( d\lambda \) is an infinitesimal increment in \( \lambda \). We will use also the unit step function \( u(z - c) \triangleq \begin{cases} 0, & z < c, \\ 1, & z \geq c \end{cases} \).
Our main result is the following observation:

**Theorem 5.1.** The duality gap remains within 1-bit and $(\Phi = [\phi^*_{m,n}(\lambda^{++})], \mathbf{p} = [p^*_{m,n}(\lambda^{++})], \mathbf{b} = [b^*_{m,n}(\lambda^{++})])$ is an optimal solution to (5.6).

This theorem indicates that, after $\lambda^*$ is obtained as described in Section 5.2.2, one may simply substitute $\lambda = \lambda^{**}$ into Section 5.2.1 to extract the jointly optimal channel pairing, power allocation, and bit loading solution. We provide the proof of Theorem 5.1 in Section 5.3.3, after presenting in Sections 5.3.1 and 5.3.2 several interesting observations and required lemmas for the proof.

### 5.3.1 Properties of $B^*(\lambda)$

As shown in Section 5.1.1, it is clear that $B^*(P_t)$ is an integer-valued increasing staircase function in the allowed total power budget $P_t$. The following lemma states a less obvious, but important fact that each step of the staircase has unit height, i.e., if $B^*(P_t) = k$ is achievable for some $P_t$ then $B^*(P_t) = i$ is achievable for some $P_t$ for all $i = 1, 2, \ldots, k - 1$. Its proof is provided in Appendix C 7.3.1.

**Lemma 5.1.** $B^*(P_t)$ is a unit-step increasing staircase function of $P_t$, such that

$$B^*(P_t) = \sum_{k} u(P_t - P_t^*(k)) . \quad (5.18)$$

Figure 5.1 illustrates the staircase function $B^*(P_t)$. Note that the corner points $(P_t^*(k), k)$ are also the points where $B^*(P_t)$ is achieved with equality for the power constraint (5.3). Further observing the line that represents $g(\lambda)$, this leads to the following lemma, whose proof is given in Appendix C 7.3.2.

**Lemma 5.2.** For all $\lambda \geq 0$, the point $(P^*(\lambda), B^*(\lambda))$ always falls on the curve $B^*(P_t)$ described by (5.6). Furthermore, for all $\lambda > 0$, $P^*(\lambda) = P_t^*(B^*(\lambda))$. 
Lemma 5.2 states that, if we substitute $P_t = P^{ts}(\lambda)$ into (5.6) and solve the optimization problem, the resultant maximum sum rate is $B^{*}(\lambda)$, and if we substitute $B = B^{*}(\lambda)$ into (5.7) and solve the optimization problem, the resultant minimum required power is $P^{ts}(\lambda)$.

We can then draw the following conclusion:

**Proposition 5.1.** $B^{*}(\lambda)$ is a unit-step decreasing staircase function of $\lambda$.

**Proof.** We need to show that $B^{*}(\lambda)$ is a decreasing staircase function, and that it does not jump down by more than one for any small increase in $\lambda$, i.e.,

$$B^{*}(\lambda) - B^{*}(\lambda + d\lambda) \leq 1 .$$

(5.19)

The first assertion can be seen considering the form of (5.10) and its geometric representation on Figure 5.1 as the highest line with slope $\lambda$ that touches a point $(P^{*}(k), k)$ on the $B^{*}(P_t)$ function. More formally, suppose there exist some $\lambda_1 < \lambda_2$ such that $B^{*}(\lambda_1) < B^{*}(\lambda_2)$. Then by Lemma 5.2 we have

$$P^{*}(\lambda_1) = P^{*}(B^{*}(\lambda_1)) < P^{*}(B^{*}(\lambda_2)) = P^{*}(\lambda_2) .$$

(5.20)
Furthermore, from (5.10) we have

\begin{align}
B^*(\lambda_2) - \lambda_1 P^*(\lambda_2) &\leq B^*(\lambda_1) - \lambda_1 P^*(\lambda_1) \quad (5.21) \\
B^*(\lambda_1) - \lambda_2 P^*(\lambda_1) &\leq B^*(\lambda_2) - \lambda_2 P^*(\lambda_2). \quad (5.22)
\end{align}

Summing the above, and re-arranging, we have

\begin{align}
(\lambda_1 - \lambda_2)(P^*(\lambda_1) - P^*(\lambda_2)) \leq 0, \quad (5.23)
\end{align}

which contradicts (5.20).

We use contradiction to show the second assertion. Suppose for some \( \lambda, k \), and \( 1 \leq i \leq k \), we have \( B^*(\lambda) = k + 1 \) and \( B^*(\lambda^+) = k - i \), where \( \lambda^+ = \lambda + d\lambda \). Consider the points \( (P^*(\lambda^+), k - i) \) and \( (P^*(\lambda), k + 1) \) on the \( P_t - B^* \) plane. By Lemma 5.2, both points are on the curve \( B^*(P_t) \), and in particular \( P_t^*(k - i) = P^*(\lambda^+) \) and \( P_t^*(k + 1) = P^*(\lambda) \).

Now consider the point \( (P_t^*(k), k) \). Lemma 5.1 ensures that such a point exists. Furthermore, we have

\begin{align}
k - \lambda P_t^*(k) &= k - \lambda^+ P_t^*(k) \\
&> k - \lambda^+ \left( \frac{i}{i + 1} P_t^*(k + 1) + \frac{1}{i + 1} P_t^*(k - i) \right) \\
&= \frac{i}{i + 1} ((k + 1) - \lambda P^*(\lambda)) \\
&\quad + \frac{1}{i + 1} ((k - i) - \lambda^+ P^*(\lambda^+)) \quad (5.24)
\end{align}

where the second line is a consequence of (5.8) by induction. This implies that either one of the following inequalities must hold,

\begin{align}
k - \lambda P_t^*(k) &> (k + 1) - \lambda P^*(\lambda), \quad (5.25) \\
k - \lambda^+ P_t^*(k) &> (k - i) - \lambda^+ P^*(\lambda^+), \quad (5.26)
\end{align}

which contradicts with the fact that \( P^*(\lambda) \) and \( P^*(\lambda^+) \) respectively maximize (5.10) for the given \( \lambda \) and \( \lambda^+ \). \qed
5.3.2 Properties of $P^*(\lambda)$

Proposition 5.1 immediately leads to the following property for $P^*(\lambda)$.

**Lemma 5.3.** $P^*(\lambda)$ is a decreasing staircase function of $\lambda$. Furthermore, $P^*(\lambda)$ and $B^*(\lambda)$ have the same discontinuity points in $\lambda$.

**Proof.** From Lemma 5.2, we have $P^*(\lambda) = P^*_t(B^*(\lambda))$. Therefore, it is constant as long as $B^*(\lambda)$ remains constant. As $\lambda$ increases, Proposition 5.1 dictates that $B^*(\lambda)$ decreases in unit steps. Each time $B^*(\lambda)$ steps down from $k$ to $k - 1$, $P^*(\lambda)$ jumps from $P^*_t(k)$ to $P^*_t(k - 1)$. Finally, Lemma 5.1 suggests that $P^*_t(k) - P^*_t(k - 1) > 0$.

Figure 5.2 illustrates the staircase function $P^*(\lambda)$. We can further characterize the relation between $P_t$ and $P^*(\lambda)$ near $\lambda = \lambda^*$ in the following:

**Proposition 5.2.** If $\lambda^* > 0$, then $\lambda^*$ is at a point where the function $P^*(\lambda)$ is discontinuous, and furthermore

$$P^*(\lambda^+) \leq P_t < P^*(\lambda^-).$$  \hspace{1cm} (5.27)

**Proof.** If there exists some $\lambda$ such that $P^*(\lambda) = P_t$, then the subgradient at $\lambda$ as defined in (5.15) is zero. Since $\lambda$ is a global minimizer of $g(\lambda)$ as long as there exists one zero subgradient at $\lambda$ [37], we have $\lambda^* = \lambda$. Then, since $P^*_t(\lambda)$ is a decreasing staircase function by Lemma 5.3, we have (5.27).

If there does not exist $\lambda$ such that $P^*(\lambda) = P_t$, then the minimum subgradient in (5.15) is

$$\theta_m \Delta \min_{\lambda} |P_t - P^*(\lambda)| > 0.$$  \hspace{1cm} (5.28)
Consider the subgradient updating algorithm to minimize \( g(\lambda) \). Suppose at iteration \( l \), we have \( P^\ast(\lambda^{(l)}) < P_t \). The Lagrange multiplier for the next iteration is updated as \( \lambda^{(l+1)} = \lambda^{(l)} + \nu^{(l)}(P^\ast(\lambda^{(l)}) - P_t) \), which results in \( \lambda^{(l+1)} < \lambda^{(l)} \). In fact, as long as \( P^\ast(\lambda^{(l)}) < P_t \), \( \lambda \) continues to decrease as \( l \) increases. Since \( \{\nu^{(l)}\} \) is a non-summable series, and \( P_t - P^\ast(\lambda^{(l)}) \geq \theta_m > 0 \) for all \( \lambda^{(l)} \), this process can continue either until \( \lambda \) decreases to 0 or until an iteration \( k > l \) such that \( P^\ast(\lambda^{(k)}) > P_t \). However, since the former case cannot occur by the proposition’s assumption that \( \lambda^* > 0 \), the latter case holds. Similarly, suppose at iteration \( l' \) we have \( P^\ast(\lambda^{(l')}) > P_t \), then there exists an iteration \( k' > l' \), such that \( P^\ast(\lambda^{(k')}) < P_t \).

The convergence of the subgradient updating algorithm, in terms of \( \min_l g(\lambda^{(l)}) \), is guaranteed [37]. This implies that \( \lambda^{(l)} \) converges to \( \lambda^* \), since there are multiple minimizers only when there exists \( \lambda \) such that \( P^\ast(\lambda) = P_t \). By the above analysis, at convergence when \( \lambda^* \) is obtained, we can have neither \( P^\ast(\lambda^*) > P_t \) nor \( P^\ast(\lambda^*) < P_t \). The only possible candidate for such \( \lambda^* \) is at a discontinuity point of the function \( P^\ast(\lambda) \) where (5.27) holds.

\[ \square \]

### 5.3.3 Proof of Main Theorem

We are now ready to prove Theorem 5.1.

**Proof.** If \( \lambda^* = 0 \), we must have \( P_t^\ast(0) \leq P_t \), since otherwise the objective in (5.11) can be decreased by increasing \( \lambda^* \) from 0. By Lemma 5.3, we have \( P_t^\ast(\lambda^{*+}) \leq P_t \), so the solution \((\Phi = [\phi_{m,n}(\lambda^{*+})], p = [p_{m,n}(\lambda^{*+})], b = [b_{m,n}(\lambda^{*+})])\) is feasible. Furthermore, we have \( B^\ast(\lambda^{*+}) = B^\ast(0) = g(\lambda^*) \) indicating that the duality gap is zero and less than 1 bit. Since \( B^\ast(0) \) maximizes the objective in (5.10), and that is the same as maximizing the objective in (5.6) when \( \lambda = 0 \), we conclude that the solution \((\Phi = [\phi_{m,n}(\lambda^{*+})], p = [p_{m,n}(\lambda^{*+})], b = [b_{m,n}(\lambda^{*+})])\) is optimal.

If \( \lambda^* > 0 \), Proposition 5.2 states that \( \lambda^* \) is at a discontinuity point of the function \( P^\ast(\lambda) \). Then, by Lemma 5.3, \( \lambda^* \) is also at a discontinuity point of the function \( B^\ast(\lambda) \). From Proposi-
tion 5.1, we see that $B^\star(\lambda)$ is a unit-step decreasing staircase function. Hence,

$$B^\star(\lambda^+) = B^\star(\lambda^-) - 1. \quad (5.29)$$

Then, we have

$$B^\star(P_t) \leq g(\lambda^+)$$

$$\leq g(\lambda^+ - \lambda^-) = B^\star(\lambda^- - (P^\star(\lambda^-) - P_t))$$

$$< B^\star(\lambda^-) = B^\star(\lambda^+) + 1, \quad (5.30)$$

where the first line is due to weak duality, the second line results from the definition of the dual function (5.10), and the third line is due to $P^\star(\lambda^-) > P_t$ from Proposition 5.2. Since $B^\star(P_t)$ is an integer, it follows that

$$B^\star(P_t) \leq B^\star(\lambda^+). \quad (5.31)$$

However, since $B^\star(\lambda^+)$ is feasible, we also have

$$B^\star(P_t) \geq B^\star(\lambda^+). \quad (5.32)$$

This implies that $B^\star(\lambda^+)$ is the maximum achievable sum rate, i.e., $(\Phi = [\phi_{m,n}(\lambda^+)])$, $p = [p_{m,n}(\lambda^+)])$, $b = [b_{m,n}(\lambda^+)])$ is an optimal solution. Moreover, by inserting $B^\star(P_t) = B^\star(\lambda^+)$ into (5.30) we have $B^\star(P_t) \leq g(\lambda^+) < B^\star(P_t) + 1$ that implies the duality gap is within 1-bit.

5.4 Complexity Reduction and Analysis

The baseline subgradient updating algorithm presented in Section 5.2.2, using the non-summable, square summable step-size rule, requires an infinite number of iterations for convergence, and presents challenges in setting a precise criterion for convergence testing. However, we can drastically reduce the computational complexity of the proposed method by terminating the subgradient updating iterations before convergence to $\lambda^*$, which still maintains our ability to
recover the optimal primal solution. A salient observation here is that, as long as \( \lambda(l) \) reaches a point where \( B^*(\lambda(l)) \) is at the same plateau as \( B^*(\lambda^{++}) \), we may stop and use \( \lambda(l) \) to extract the optimal primal solution.

The following theorem formalizes this idea. It gives two conditions upon which the algorithm may be safely terminated. Figure 5.2 contains a pictorial illustration of the early termination conditions.

**Theorem 5.2.** At iterations \( l \) and \( l + 1 \), if

1. \(|B^*(\lambda(l)) - B^*(\lambda(l+1))| = 1\), and
2. \( \text{sgn} \left( P_t - P^*(\lambda(l)) \right) \neq \text{sgn} \left( P_t - P^*(\lambda(l+1)) \right) \),

then \( B^*(P_t) = \min \{ B^*(\lambda(l)), B^*(\lambda(l+1)) \} \).

**Proof.** Suppose \( B^*(\lambda(l)) > B^*(\lambda(l+1)) \). This gives rise to \( P^*(\lambda(l)) > P^*(\lambda(l+1)) \), since both \( B^* \) and \( P^* \) are decreasing in \( \lambda \). Thus, condition 2) implies \( P^*(\lambda(l+1)) < P_t < P^*(\lambda(l)) \).

Similar to (5.30), we have

\[
B^*(P_t) < B^*(\lambda(l)).
\] (5.33)

Since \( B^*(P_t) \) is an integer, it follows that

\[
B^*(P_t) \leq B^*(\lambda(l)) - 1 = B^*(\lambda(l+1))
\] (5.34)

where the equality is due to condition 1) above. Since \( B^*(\lambda(l+1)) \) is also a feasible solution with \( P^*(\lambda(l+1)) < P_t \), (5.34) implies that \( B^*(P_t) = B^*(\lambda(l+1)) \).

For the case where \( B^*(\lambda(l)) < B^*(\lambda(l+1)) \), a similar argument applies by swapping index \( (l+1) \) for \( (l) \), to show that \( B^*(P_t) = B^*(\lambda(l)) \). \( \Box \)

From this, we see that there is no need for the subgradient updating algorithm to converge exactly to \( \lambda^* \). Instead, as long as \( \lambda(l) \) reaches within a certain neighborhood of \( \lambda^* \), which is determined by the plateau defined by \( B^*(\lambda^{++}) = B^*(P_t) \), the optimal primal solution can be
extracted. This enables a variant of the proposed algorithm that can be shown to require only polynomial complexity in the number of channels $N$.

We first make the following observation that $\lambda^*$ is bounded by a constant.

**Lemma 5.4.** $\lambda^*$ is upper bounded by a constant $\lambda_{\text{max}}$ independent of $N$.

**Proof.** Consider $\lambda^{++}$ instead of $\lambda^*$ to avoid ambiguity at the discontinuities of $B^*(\lambda)$ and $P^*(\lambda)$. Let $k = B^*(\lambda^{++})$. Then, by (5.10) and Lemma 5.2, we have

$$k - \lambda^{++} P^*_t(k) \geq (k - 1) - \lambda^{++} P^*_t(k - 1).$$

This implies that

$$\lambda^* \leq \frac{1}{P^*_t(k) - P^*_t(k - 1)} \leq \frac{1}{P^*_t(1) - P^*_t(0)} = \frac{1}{P^*_t(1)} = \frac{1}{\sum_{m,n} \phi^*_m p^*_m},$$

where the second line is due to (5.8), and $\phi^*_m$ and $p^*_m$ denote the optimal solution to (5.7) that gives $P^*_t(1)$. Since the sum rate is 1, all power in $P^*_t(1)$ must go to the one path $(m', n')$ that supports $b_{m'n'} = 1$. It follows that

$$\lambda^* \leq \frac{1}{p^*_{m'n'}}.$$

Therefore, $\lambda^*$ is upper bounded by the minimum power required to transmit at bit rate $b_{mn} = 1$ for any path $(m, n)$, which we denote by $\lambda_{\text{max}}$. Since such minimum required power is generally determined by the modulation and coding schemes, it is independent of the number of channels $N$.

Next, we may modify the original subgradient updating in Section 5.2.2 to the following without affecting the optimality of the overall method:

$$\lambda^{(l+1)} = \min\{\max\{\lambda^{(l)} - \theta(\lambda^{(l)}) \nu^{(l)}, 0\}, \lambda_{\text{max}}\}.$$

Furthermore, without the need for exact convergence to $\lambda^*$, we may adopt a constant step-size rule [38], i.e., $\nu^{(l)} = \nu$, which improves the convergence speed when the iteration count...
is high. The polynomial-time complexity of this alternate method is stated in the following theorem:

**Theorem 5.3.** To achieve within an arbitrary \( \epsilon > 0 \) neighborhood of the optimal \( g(\lambda^*) \), using a constant step-size in subgradient updating, the alternate method with (5.38) has polynomial computational complexity in \( N \).

**Proof.** At each iteration of the standard subgradient updating algorithm, the procedures described in Section 5.2.1 have overall computational complexity polynomial in \( N \), since the two-dimensional assignment problem can be solved in \( O(N^3) \). Therefore, it remains to show that the total number of iterations is polynomial in \( N \).

At each iteration \( l \), the subgradient is

\[
\theta(\lambda(l)) = P_t - P^*(\lambda(l)) \leq P_t.
\] (5.39)

Furthermore, \( P^*(\lambda(l)) \) is upper bounded by the total power to support transmission at the maximum data rate \( M \) for each of the \( N \) chosen paths, which has order \( O(N) \). Hence, \( |\theta(\lambda(l))| \) is upper bounded by \( \Theta = O(N) \).

If we choose \( \lambda(0) \) in the interval \([0, \lambda_{max}]\), then the distance between \( \lambda(0) \) and \( \lambda^* \) is upper bounded by \( \lambda_{max} \). Then, it can be shown that, at the \( l \)th iteration, the distance between the current best objective to the optimum objective \( g(\lambda^*) \) is upper bounded, by \( \frac{\lambda^2_{max} + \nu^2\Theta^2 l}{2\nu l} \) if a constant step-size \( \nu(l) = \nu \) is used [38]. If we set \( \nu = \epsilon/\Theta^2 \), it is easy to see that the above bound is less than \( \epsilon \) when \( l > \frac{\lambda^2_{max} \Theta^2}{\epsilon^2} = O(N^2) \). Hence, the number of required iterations until convergence is \( O(N^2) \).

Most mixed integer optimization problems are NP-hard, and in particular, many previously proposed solutions to optimize resource allocation with bit loading are suboptimal. Therefore, it is notable to observe that for the problem under consideration, one can achieve jointly optimal solutions using the proposed method within polynomial time in \( N \). In the context of practical OFDM systems such as IEEE 802.16, which operate over hundreds or thousands of subcarri-
ers, the proposed method provides substantial savings in computing resources over brute-force search methods.

5.5 Numerical Results

This section presents simulation results to assess and compare the performance of the jointly optimal scheme with that of the following suboptimal schemes:

- Separate Optimization: The two-stage solution proposed in [82], which first pairs the channels based on their sorted channel gains and then allocates bits over the paired channels via the greedy algorithm in [70].

- No Pairing: Carry out bit loading and power allocation without pairing (i.e., same channel is used for relaying a received signal from that channel).

- Uniform PA: Allocate available power uniformly among the paths without pairing.

We consider a multi-channel AF relaying system where a relay lies between a pair of source and destination with $d_{sr} = d_{rd} = d$, where $d_{sr}$ and $d_{rd}$ denote the source-relay and relay-destination distances, respectively. A 4-tap frequency selective propagation channel is assumed for each hop. To normalize the effect of power over different network settings, we define the average received SNR under a hypothetical uniform allocation of the total transmission power, which for the S-R link is $\text{SNR}_{avg}^{S-R} = \frac{P_t^s}{N \sigma^2 d_{sr}^\kappa}$ and for the R-D link is $\text{SNR}_{avg}^{R-D} = \frac{P_t^r}{N \sigma^2 d_{rd}^\kappa}$, where $P_t^s$ and $P_t^r$ are the total source and relay available powers, respectively, and $\kappa = 3$ denotes the pathless exponent.

In this example, we focus on optimal resource allocation at the relay. Hence, $P_t$ and $p_{mn}$ in our analysis here refers to $P_t^r$ and the power allocated by the relay to path $(m, n)$, respectively. The SNR gap $\Gamma$ is set to 1. Then the total achievable rate function $R_{mn}(p_{mn}) = \frac{1}{2} \log(1 + \frac{p_{mn} a_m p_{mn} b_n}{1 + p_m a_m + p_{mn} b_n})$, where $p_m^s$ is a given power allocated to the S-R channel $m$. In the following, we assume uniform power allocation at the source across all S-R channels, so $p_m^s = \frac{P_s}{N}$, although
other arbitrary power allocation is applicable as well. The maximum bit rate per channel $M$ is set to 5.

Our first experiment compares the performance of various schemes at different $\text{SNR}_{\text{avg}}^{R-D}$ levels for $N = 16$. Fig. 5.3 depicts the average bit rate per channel vs. $\text{SNR}_{\text{avg}}^{R-D}$, while $\text{SNR}_{\text{avg}}^{S-R}$ is fixed to 10 dB. As we observe, the jointly optimal scheme outperforms the other suboptimal solutions, offering as much as 1dB improvement over the Separate Optimization scheme for a large range of SNR values. Furthermore, the only difference between Separate Optimization and No Pairing is that the former scheme conducts pairing while the latter does not. It is therefore interesting to notice the substantial gain provided by channel pairing alone. At high received $\text{SNR}_{\text{avg}}^{R-D}$, it is anticipated that the performance of all schemes converge to a single point as the S-R link becomes the bottleneck of communication.

The second experiment examines the effect of the number of channels on the performance of various resource allocation schemes. This is done under the same level of channel gain variation across channels, and under the fixed channel spacing in terms of bandwidth. Fig. 5.4 illustrates the average bit rate per channel vs. the number of channels for $\text{SNR}_{\text{avg}}^{S-R} = \text{SNR}_{\text{avg}}^{R-D} = 10$ dB. We observe that the gap between different schemes widens as the number of channels increase. This indicates that as more channels become available, they can be exploited more judiciously for pairing and power optimization.

5.6 Summary

This chapter has tackled the problem of jointly optimal CP, PA, and bit loading in a dual-hop multi-channel system. Despite the non-zero duality gap due to its mixed-integer nature, it is shown that the problem’s special structure allows the extraction of an exact optimal solution. The proposed approach is applicable to general relaying strategies, modulation schemes, and performance metrics. Some complexity reduction techniques are developed and the algorithm runtime are demonstrated to be polynomial in the number of channels.
Figure 5.3: Per channel rate vs. SNR$_{avg}^{R-D}$ for multi-channel AF relaying with SNR$_{avg}^{S-R} = 10dB$.

Figure 5.4: Per channel rate vs. number of channels for multi-channel AF relaying with SNR$_{avg}^{S-R} = \text{SNR}_{avg}^{R-D} = 10dB$. 
Chapter 6

Conclusion

This thesis has investigated the problem of optimal resource management for a variety of multi-channel relaying architectures to be used in the emerging wireless systems. For each architecture, we have provided the optimal and viable resource management solution to maximize the overall system performance and throughput. We have adopted the channel pairing (CP) capability that allows the relay to receive a signal from one channel and transmit a processed version of the signal on a different channel.

In Chapter 3, we considered a multi-hop multi-channel setup for which we addressed the problem of channel pairing jointly with power allocation (PA), which determines each channel’s power across the channels. Despite the correlation between these two problems, we proved that the jointly optimal solution can be decoupled into two separate CP optimization and PA optimization, thus significantly reducing the search complexity. We further showed that the CP problem can be again decomposed into independent pairing problems at each relay based on sorted channel gains, independent of PA on each channel. This resulted in simple and distributed implementation of the relay system for optimal operation. Through a dual-decomposition approach, we also provided the solution for optimizing PA for both DF and AF relaying.

We then in Chapter 4 focused on a multi-user scenario where we specifically tackled the
problem of channel-user assignment together with CP and PA for multi-channel relaying communication. We argued that the problem typically arises in cellular communication or wireless local area networks, through either dedicated relay stations or users temporarily serving as relay nodes. Despite being formulated as a mixed integer programming (IP), we show that there is an efficient method to jointly optimize channel pairing, channel-user assignment, and power allocation in such general dual-hop relaying networks. The proposed solution framework relied on continuous relaxation and Lagrangian dual minimization. Although this approach generally provides only heuristic or approximate solutions for an IP problem, by exploring the rich structure in our problem, we show that judicious reformulation and choices of the optimization trajectory can preserve both the binary constraints and the strong Lagrangian duality property of the continuous version (i.e., zero duality gap), thus enabling a jointly optimal solution.

Chapter 5 focused on the bit loading problem along with PA and CP problems in a dual-hop multi-channel relaying system with the aim of maximizing the overall bit rate. The resulting optimization problem is of non-convex mixed IP in nature, with both a discrete objective and a discrete feasible region. The Lagrange duality approach can be applied to such problems, but it generally does not lead to an optimal solution due to the non-zero duality gap. Nonetheless, we show that by exploring the structure of our problem, we are able to bound the gap to the original objective to be within one bit. This knowledge allows us to extract the exact optimal integer solution. We further develop numerical techniques to reduce the algorithm runtime. This suggests that it is more amenable to practical implementation than combinatoric search approaches, especially for practical systems, such as IEEE 802.16 OFDMA, with hundreds to thousands of frequency sub-channels.
Chapter 7

Appendices

7.1 Appendix A

7.1.1 Proof of Proposition 3.1

At relay 1, there are two ways to pair the channels: (1) channels 1 and 2 over hop 1 are matched with channels 1 and 2 over hop 2, respectively; (2) channels 1 and 2 over hop 1 are matched with channels 2 and 1 over hop 2, respectively. These two ways of pairing lead to the following two sets of disjoint paths from the source to the destination: \( \{ P_{1}^{(1)} \} = \{ (1, 1, c(3, 1)), (2, 2, c(3, 2)) \} \) and \( \{ P_{1}^{(2)} \} = \{ (1, 2, c(3, 1)), (2, 1, c(3, 2)) \} \), where the superscript \( j \) in \( \{ P_{1}^{(j)} \} \) indicates a different set of path selection.

By considering the equivalent channels from the source to the second relay, using the existing optimality result for dual-hop relaying [62], it is easy to see that \( c(3, 1) = 1 \) and \( c(3, 2) = 2 \) are optimal for \( \{ P_{1}^{(1)} \} \). Furthermore, we only need to show

\[
\log_{2} \left( 1 + \gamma_{SD}(P_{1}^{(1)}) \right) + \log_{2} \left( 1 + \gamma_{SD}(P_{2}^{(1)}) \right) \geq \log_{2} \left( 1 + \gamma_{SD}(P_{1}^{(2)}) \right) + \log_{2} \left( 1 + \gamma_{SD}(P_{2}^{(2)}) \right), \tag{7.1}
\]

for the case of \( c(3, 1) = 1 \) and \( c(3, 2) = 2 \) for both \( \{ P_{1}^{(1)} \} \) and \( \{ P_{1}^{(2)} \} \), since the case of \( c(3, 1) = 2 \) and \( c(3, 2) = 1 \) for \( \{ P_{1}^{(2)} \} \) can be similarly proven. Inequality (7.1) for the AF and
DF relaying cases are separately proven as follows:

**AF Relaying**  By inserting (3.1) into inequality (7.1) we need to show

\[
\left(1 + (Q_1^{(1)} - 1)^{-1}\right)\left(1 + (Q_2^{(1)} - 1)^{-1}\right) \geq \\
\left(1 + (Q_1^{(2)} - 1)^{-1}\right)\left(1 + (Q_2^{(2)} - 1)^{-1}\right),
\]  \(7.2\)

where

\[
Q_1^{(1)} = \left(1 + \frac{1}{\gamma_{1,1}}\right)\left(1 + \frac{1}{\gamma_{2,1}}\right)\left(1 + \frac{1}{\gamma_{3,1}}\right),
\]

\[
Q_2^{(1)} = \left(1 + \frac{1}{\gamma_{1,2}}\right)\left(1 + \frac{1}{\gamma_{2,2}}\right)\left(1 + \frac{1}{\gamma_{3,2}}\right),
\]

\[
Q_1^{(2)} = \left(1 + \frac{1}{\gamma_{1,1}}\right)\left(1 + \frac{1}{\gamma_{2,2}}\right)\left(1 + \frac{1}{\gamma_{3,1}}\right),
\]

\[
Q_2^{(2)} = \left(1 + \frac{1}{\gamma_{1,2}}\right)\left(1 + \frac{1}{\gamma_{2,1}}\right)\left(1 + \frac{1}{\gamma_{3,2}}\right).
\]  \(7.3\)

The following lemma is used to prove (7.2)

**Lemma 7.1.**  With condition (3.10), we have

\[
(Q_1^{(1)} - 1)(Q_2^{(1)} - 1) \leq (Q_1^{(2)} - 1)(Q_2^{(2)} - 1).
\]  \(7.4\)
Proof. By substituting (7.3) in the following term and expanding it, we have

\[(Q_1^{(1)} - 1)(Q_2^{(1)} - 1) - (Q_1^{(2)} - 1)(Q_2^{(2)} - 1)\]

\[= Q_1^{(2)} + Q_2^{(2)} - Q_1^{(1)} - Q_2^{(1)} \quad (7.5)\]

\[= \left(1 + \frac{1}{\gamma_{1,1}}\right)\left(1 + \frac{1}{\gamma_{2,2}}\right)\left(1 + \frac{1}{\gamma_{3,1}}\right) + \left(1 + \frac{1}{\gamma_{1,2}}\right)\left(1 + \frac{1}{\gamma_{2,1}}\right)\left(1 + \frac{1}{\gamma_{3,2}}\right) - \left(1 + \frac{1}{\gamma_{1,1}}\right)\left(1 + \frac{1}{\gamma_{2,1}}\right)\left(1 + \frac{1}{\gamma_{3,1}}\right) - \left(1 + \frac{1}{\gamma_{1,2}}\right)\left(1 + \frac{1}{\gamma_{2,2}}\right)\left(1 + \frac{1}{\gamma_{3,2}}\right)\]

\[= \left(\frac{1}{\gamma_{2,2}} - \frac{1}{\gamma_{2,1}}\right) \times \left(\left(1 + \frac{1}{\gamma_{3,2}}\right)\left(1 + \frac{1}{\gamma_{1,2}}\right) - \left(1 + \frac{1}{\gamma_{1,1}}\right)\left(1 + \frac{1}{\gamma_{3,1}}\right)\right) \quad (7.6)\]

\[\leq 0\]

where we have used the fact that \(Q_1^{(1)}Q_2^{(1)} = Q_1^{(2)}Q_2^{(2)}\) to arrive at (7.5). From condition (3.10), the first product term in (7.6) is negative and the second product term is positive, and therefore we obtain the last inequality.

Consider the subtraction of the RHS from the LHS of (7.2),

\[\text{LHS of (7.2) - RHS of (7.2)}\]

\[= \left(\frac{(Q_1^{(1)} - 1)^{-1} + (Q_2^{(1)} - 1)^{-1} + (Q_1^{(1)} - 1)^{-1}(Q_2^{(1)} - 1)^{-1}}{A}\right) - \left(\frac{(Q_1^{(2)} - 1)^{-1} + (Q_2^{(2)} - 1)^{-1} + (Q_1^{(2)} - 1)^{-1}(Q_2^{(2)} - 1)^{-1}}{B}\right)\]

\[\geq A(Q_1^{(1)} - 1)(Q_2^{(1)} - 1) - B(Q_1^{(2)} - 1)(Q_2^{(2)} - 1) \quad (7.7)\]

\[= Q_2^{(1)} + Q_1^{(1)} - Q_1^{(2)} - Q_2^{(2)}\]

\[= \left(\frac{1}{\gamma_{2,2}} - \frac{1}{\gamma_{2,1}}\right)\left(\left(1 + \frac{1}{\gamma_{1,2}}\right)\left(1 + \frac{1}{\gamma_{3,2}}\right) - \left(1 + \frac{1}{\gamma_{1,1}}\right)\left(1 + \frac{1}{\gamma_{3,1}}\right)\right) \quad (7.8)\]

\[\geq 0,\]
where the inequality (7.7) holds because of Lemma 7.1, and the fact that $Q^{(j)}_i - 1 > 0$, for $i = 1, 2$ and $j = 1, 2, 3$; and the inequality (7.8) holds because of condition (3.10).

**DF Relaying** Inserting (3.3) into inequality (7.1), we need to show

$$
(1 + \min(\gamma_{1,1}, \gamma_{2,1}, \gamma_{3,1}))(1 + \min(\gamma_{1,2}, \gamma_{2,2}, \gamma_{3,2})) \geq
(1 + \min(\gamma_{1,1}, \gamma_{2,2}, \gamma_{3,1}))(1 + \min(\gamma_{1,2}, \gamma_{2,1}, \gamma_{3,2})).
$$

(7.9)

We can verify (7.9) by enumerating all possible relations among $\gamma_{m,n}$, for all $m = 1, 2, 3$ and $n = 1, 2$, subject to condition (3.10). For example, when $\gamma_{1,1} \leq \gamma_{2,1} \leq \gamma_{3,1}$, $\gamma_{1,2} \leq \gamma_{2,2} \leq \gamma_{3,2}$, $\gamma_{2,2} \leq \gamma_{1,1} \leq \gamma_{3,2}$, and $\gamma_{3,2} \leq \gamma_{2,1}$, (7.9) reduces to

$$(1 + \gamma_{1,1})(1 + \gamma_{1,2}) \geq (1 + \gamma_{2,2})(1 + \gamma_{1,2}).$$

The above inequality clearly holds based on the assumption of $\{\gamma_{i,j}\}$ relations. Inequality (7.9) can be similarly verified for all other $\{\gamma_{i,j}\}$ relations. The details are omitted for brevity.

### 7.1.2 Proof of Lemma 3.1

Let $R_n$ denote the end-to-end data rate on path $n$, we have $R_n = \frac{1}{F_s} \log_2(1 + \gamma_n)$, $n = 1, \ldots, N$, where

$$
\gamma_n = \left( \sum_{m=1}^{M} \frac{1}{P_{m,n} a_{m,n}} \right)^{-1}.
$$

(7.10)

Then $R^{up}_t = \sum_{n=1}^{N} R_n$. To show $R^{up}_t$ is concave in $\{P_{m,n}\}$, it suffices to show that each $R_n$ is concave in $\{P_{mn}\}$. The concavity proof of $R_n$ follows the concavity of $\gamma_n$ due to the composition rules which preserve concavity [33]. For simplicity, we drop the subscript $n$ from notations in (7.10). In the following we prove that $\gamma(\bar{P})$ is concave in $\bar{P}$, where $\bar{P} = [P_1, \ldots, P_M]^T$. The second-order partial derivatives of $\gamma(\bar{P})$ are given by

$$
\frac{\partial^2 \gamma(\bar{P})}{\partial P^2_j} = -\frac{2}{a_j P_j^3} \left( \sum_{i=1}^{M} \frac{1}{a_i P_i} \right)^{-2} + \frac{2}{a_j^2 P_j^4} \left( \sum_{i=1}^{M} \frac{1}{a_i P_i} \right)^{-3}
$$

(7.11)
and
\[
\frac{\partial^2 \gamma (\bar{P})}{\partial P_j \partial P_k} = \frac{1}{a_j P_{j}^2} \frac{2}{a_k P_{k}^2} \left( \sum_{i=1}^{M} \frac{1}{a_i P_{i}} \right)^{-3}, \quad \text{for } k \neq j. \tag{7.12}
\]

Hence, the Hessian matrix $\Delta^2 \gamma (\bar{P})$ can be expressed as
\[
\Delta^2 \gamma (\bar{P}) = \left( \sum_{i=1}^{M} \frac{1}{a_i P_{i}} \right)^{-3} \left( -2 \left( \sum_{i=1}^{M} \frac{1}{a_i P_{i}} \right) \text{diag} \left( \frac{1}{a_1 P_{1}^3}, \ldots, \frac{1}{a_M P_{M}^3} \right) + 2qq^T \right), \tag{7.13}
\]
where $q = [q_1, \ldots, q_M]^T$ with $q_m = \frac{1}{a_m P_{m}^2}$, $m = 1, \ldots, M$, and $\text{diag}(x)$ denotes the diagonal matrix with diagonal elements being the elements in vector $x$. To prove concavity, we need to show $\Delta^2 \gamma (\bar{P}) \succeq 0$. For any vector $v = [v_1, \ldots, v_M]^T$, we have
\[
v^T \Delta^2 \gamma (\bar{P}) v = 2 \left( \sum_{i=1}^{M} \frac{1}{a_i P_{i}} \right)^{-3} \left( - \left( \sum_{i=1}^{M} \frac{1}{a_i P_{i}} \right) \sum_{i=1}^{M} \frac{1}{a_i P_{i}^3} v_i^2 + \left( \sum_{i=1}^{M} \frac{v_i}{a_i P_{i}^2} \right)^2 \right) \leq 0
\]
where the inequality is obtained by using the Cauchy-Schwarz inequality $(e^T e)(e^T c) \geq (e^T c)^2$ for two vectors $e = [e_1, \ldots, e_M]^T$ and $c = [c_1, \ldots, c_M]^T$, with $e_i = \frac{1}{\sqrt{a_i P_{i}^3}}$ and $c_i = \frac{v_i}{\sqrt{a_i P_{i}^3}}$. Therefore, $\Delta^2 \gamma (\bar{P}) \succeq 0$.

### 7.2 Appendix B

#### 7.2.1 Derivation of Equation (4.18)

For notational simplicity, we drop all subscripts $m$, $n$, and $k$ from (5.12). We have the following maximization problem, which can be solved in the two cases below.
\[
\max_{P^s, P^r} \frac{w}{2} \phi \min \left\{ \log \left( 1 + \frac{a P^s}{\phi} \right), \log \left( 1 + \frac{c P^s}{\phi} + \frac{b P^r}{\phi} \right) \right\} - (\lambda_s + \lambda_t) P^s - (\lambda_r + \lambda_t) P^r \tag{7.14}
\]
\[
s.t. \quad P^s, P^r \geq 0.
\]
Case One: $a \leq c$

In this case, the first term inside the $\min$ function in (7.14) is always smaller than the second term. Hence, (7.14) is reduced to

$$\max_{P^s, P^r} \frac{w}{2} \tilde{\phi} \log \left( 1 + \frac{a P^s}{\tilde{\phi}} \right) - (\lambda_s + \lambda_t) P^s - (\lambda_r + \lambda_t) P^r$$

s.t. $P^s, P^r \geq 0$. \hfill (7.15)

Then, the optimal solutions from water-filling are obtained as

$$P^{s*} = \left( \frac{w}{2(\lambda_s + \lambda_t) \ln 2} \frac{1}{a} \right)^{+} \tilde{\phi}, \quad P^{r*} = 0$$ \hfill (7.16)

Case Two: $a > c$

For this more complicated case, we propose the following solution. We inspect the two possible outcomes in comparing the first and second terms in the $\min$ function in (7.14) at optimality. Two separate maximization of (7.14) are performed under the constraint of either outcome. Then, the optimal $(P^s, P^r)$ is given by the better of these two solutions.

Assumption 1: $a P^{s*} \leq b P^{r*} + c P^{s*}$: Under this assumption, we have $b > 0$ and the following optimization problem:

$$\max_{P^s, P^r} \frac{w}{2} \tilde{\phi} \log \left( 1 + \frac{a P^s}{\tilde{\phi}} \right) - (\lambda_s + \lambda_t) P^s - (\lambda_r + \lambda_t) P^r$$

s.t. (i) $P^s, P^r \geq 0$

(ii) $a P^s \leq b P^r + c P^s$. \hfill (7.17)

It has two possible solutions from the KKT conditions. One is obtained when the Lagrange multiplier corresponding to constraint (ii) is zero and the constraint is strictly satisfied. This implies that

$$P^{s*} = \left( \frac{w}{2(\lambda_s + \lambda_t) \ln 2} \frac{1}{a} \right)^{+} \tilde{\phi}, \quad P^{r*} = 0.$$ \hfill (7.18)

However, this solution contradicts with the assumption that (ii) is strictly satisfied. The other, correct solution occurs at the border $P^r = \frac{a-c}{b} P^s$. By inserting this into the objective function,
we have

\[ P^{ss} = \left( \frac{wb}{2b(\lambda_s + \lambda_t)\ln 2 + 2(a - c)(\lambda_r + \lambda_t)\ln 2} - \frac{1}{a} \right)^+ \phi, \quad (7.19) \]

\[ P^{rs} = \frac{a - c}{b} P^{ss}. \]

**Assumption 2:** \( aP^{ss} \geq bP^{rs} + cP^{ss} \): Under this assumption, we have the following optimization problem:

\[
\begin{align*}
\max_{P^s, P^r} \frac{w}{2} \phi \log \left( 1 + \frac{cP^s}{\phi} + \frac{bP^r}{\phi} \right) - (\lambda_s + \lambda_t)P^s - (\lambda_r + \lambda_t)P^r \\
s.t. \quad (i) \ P^s, P^r \geq 0 \\
\quad (ii) \ aP^s \geq bP^r + cP^s
\end{align*}
\]

(7.20)

From the KKT conditions, at optimality, either \( bP^{rs} = (a - c)P^{ss}, \) or the Lagrange multiplier corresponding to constraint (ii) is zero and the constraint is strictly satisfied.

In the former case, \( b > 0 \) since \( a \neq c. \) Furthermore, since the first and second terms in the min function in (7.14) are the same, we obtain the same solution as in (7.19).

In the latter case, we define two new variables \( V^s = (\lambda_s + \lambda_t)P^s \) and \( V^r = (\lambda_r + \lambda_t)P^r. \)

Substituting them into the objective of (7.20), we have

\[
\begin{align*}
\max_{V^s, V^r} \frac{w}{2} \phi \log \left( 1 + \frac{cV^s}{\phi(\lambda_s + \lambda_t)} + \frac{bV^r}{\phi(\lambda_r + \lambda_t)} \right) - V^s - V^r.
\end{align*}
\]

(7.21)

The solution depends on the relation between \( \frac{c}{\lambda_s + \lambda_t} \) and \( \frac{b}{\lambda_r + \lambda_t} \):

- If \( \frac{c}{\lambda_s + \lambda_t} > \frac{b}{\lambda_r + \lambda_t}, \) then we have \( V^{rs} = 0, \) since otherwise a better solution to (7.21) would be \( (V^s = V^{ss} + V^{rs}, V^r = 0). \) Substituting \( V^{rs} = 0 \) into (7.21), we have
  \[ V^{ss} = \left[ \frac{w}{2\ln 2} - \frac{\lambda_s + \lambda_t}{c} \right]^+ \phi. \]

- If \( \frac{c}{\lambda_s + \lambda_t} = \frac{b}{\lambda_r + \lambda_t}, \) (7.21) is a function of \( (V^s + V^r) \) only, and \( V^{rs} = 0 \) is a maximizer.
  
  Hence, again we have \( V^{ss} = \left[ \frac{w}{2\ln 2} - \frac{\lambda_s + \lambda_t}{c} \right]^+ \phi. \)

- If \( \frac{c}{\lambda_s + \lambda_t} < \frac{b}{\lambda_r + \lambda_t}, \) similarly we have \( V^{ss} = 0. \) However, this together with our assumption that constraint (ii) of (7.20) is strictly satisfied, i.e., \( bP^{rs} < (a - c)P^{ss}, \) implies

that \( V^{**} < 0 \), which is not a feasible solution. Therefore, in this case at optimality the condition \( bP^{**} = (a - c)P^{**} \) prevails.

### 7.2.2 Proof of Lemma 4.2

We note that there exists at least one index vector \((m', n', k')\) such that \( \phi^*_{m' n' k'}(\lambda^*) = 1 \). Furthermore, this chosen path must have non-degenerate user weight and channel gains so that the weighted rate function \( w_{k'} R(m', n', k') \) is not uniformly zero, i.e., \( w_{k'} > 0 \), \( a_{m'} > 0 \), and \( b_{n' k'} + c_{m' k'} > 0 \).

Suppose there exists \( \phi^*_{m' n' k'}(\lambda^*) = 1 \) such that \( P^{**}_{m' n' k'}(\lambda^*) \) is either \( \left[ \frac{w_{k'}}{\alpha(\lambda^*_s + \lambda^*_t)} - \frac{1}{a_{m'}} \right]^+ \) or \( \left[ \frac{w_{k'}}{\alpha(\lambda^*_s + \lambda^*_t)} - \frac{1}{c_{m' k'}} \right]^+ \). In the former case, \( a_{m'} \leq c_{m' k'} \), and the latter, \( a_{m'} > c_{m' k'} \) and \( \frac{c_{m' k'}}{\lambda^*_t + \lambda^*_s} \geq \frac{b_{n' k'}}{\lambda^*_t + \lambda^*_s} \). Furthermore, since \( b_{n' k'} \) and \( c_{m' k'} \) cannot both be zero in the latter case, \( c_{m' k'} > 0 \). Then we have

\[
\min \{ P_s, P_t \} \geq \sum_{m, n, k} P^{**}_{mnk}(\lambda^*) \geq P^{**}_{m'n'k'}(\lambda^*)
\]

\[
\geq \min\left\{ \left[ \frac{w_{k'}}{\alpha(\lambda^*_s + \lambda^*_t)} - \frac{1}{a_{m'}} \right]^+, \left[ \frac{w_{k'}}{\alpha(\lambda^*_s + \lambda^*_t)} - \frac{1}{c_{m' k'}} \right]^+ \right\}.
\]

\[
\geq \frac{w_{k'}}{\alpha(\lambda^*_s + \lambda^*_t)} - \min\{a_{m'}, c_{m' k'}\}.
\]

(7.22)

Hence, \( \lambda^*_s + \lambda^*_t \geq \frac{w_{k'}}{\alpha \min\{a_{m'}, c_{m' k'}\}} \), so \( \lambda^* \in R_1 \).

Otherwise, for all \( \phi^*_{mnk}(\lambda^*) = 1 \), we have \( a_m > c_{mk} \) and \( P^{**}_{mnk}(\lambda^*) = \left[ \frac{w_{k'} b_{nk}}{\alpha(b_{nk}(\lambda^*_s + \lambda^*_t) + (a_m - c_{mk})(\lambda^*_s + \lambda^*_t))} - \frac{1}{a_m} \right]^+ \). We proceed with the following cases:

- If there exists \( \phi^*_{m'n'k'}(\lambda^*) = 1 \) such that \( c_{m' k'} > 0 \) and \( \frac{c_{m' k'}}{\lambda^*_t + \lambda^*_s} < \frac{b_{n' k'}}{\lambda^*_t + \lambda^*_s} \), then we have

\[
\min \{ P_s, P_t \} \geq \sum_{m, n, k} P^{**}_{mnk}(\lambda^*) \geq P^{**}_{m'n'k'}(\lambda^*)
\]

\[
\geq \frac{w_{k'} b_{n' k'}}{\alpha(b_{n' k'}(\lambda^*_s + \lambda^*_t) + (a_{m'} - c_{m' k'})(\lambda^*_s + \lambda^*_t))} - \frac{1}{a_{m'}},
\]

\[
\geq \frac{w_{k'} b_{n' k'}}{\alpha(b_{n' k'}(\lambda^*_s + \lambda^*_t) + \frac{b_{n' k'}(a_{m'} - c_{m' k'})}{c_{m' k'}}(\lambda^*_s + \lambda^*_t))} - \frac{1}{a_{m'}},
\]

(7.23)

which implies that \( \lambda^*_s + \lambda^*_t \geq \frac{w_{k'} c_{m' k'}}{\alpha(a_m \min\{P_s, P_t\} + 1)} \), and hence \( \lambda^* \in R_1 \).
- Else, if there exists \( \phi^*_m \) such that \( \frac{c_{m'k'}}{\lambda_s + \lambda_t} \geq \frac{b_{n'k'}}{\lambda_s + \lambda_t} \), then \( c_{m'k'} > 0 \) since \( b_{n'k'} \) and \( c_{m'k'} \) cannot both be zero. In this case, the third expression in (4.18) applies to the path \((m', n', k')\). Since

\[
\left[ \frac{w_{k'}}{\alpha(b_{n'k'}(\lambda_s^* + \lambda_t^*) + (a_m - c_{m'k'})(\lambda_s^* + \lambda_t^*))} - \frac{1}{a_{m'}} \right]^+ \text{ is the optimal PA for this path, we have}
\]

\[
\mathcal{L}_{m'n'k'}(1, P_{m'n'k'}^{s1}, P_{m'n'k'}^{r1}, \lambda^*) \geq \mathcal{L}_{m'n'k'}(1, P_{m'n'k'}^{s2}, P_{m'n'k'}^{r2}, \lambda^*),
\]

where

\[
P_{m'n'k'}^{s1} \equiv \left[ \frac{w_{k'}b_{n'k'}}{\alpha(b_{n'k'}(\lambda_s^* + \lambda_t^*) + (a_m - c_{m'k'})(\lambda_s^* + \lambda_t^*))} - \frac{1}{a_{m'}} \right]^+
\]

\[
P_{m'n'k'}^{s2} \equiv \left[ \frac{w_{k'}}{\alpha(\lambda_s^* + \lambda_t^*)} - \frac{1}{c_{m'k'}} \right]^+
\]

(7.25)

correspond to the cases \( a_{m'} P_{m'n'k'}^{s1} = b_{n'k'} P_{m'n'k'}^{r1} + c_{m'k'} P_{m'n'k'}^{s1} \) and \( a_{m'} P_{m'n'k'}^{s2} > b_{n'k'} P_{m'n'k'}^{r2} + c_{m'k'} P_{m'n'k'}^{s2} \), respectively. This implies that

\[
\frac{w_{k'}}{2} \log(1 + a_{m'} P_{m'n'k'}^{s1}) - (\lambda_s^* + \lambda_t^*)(1 + \frac{a_{m'} - c_{m'k'}}{b_{n'k'}}) P_{m'n'k'}^{s1} \geq \frac{w_{k'}}{2} \log(1 + c_{m'k'} P_{m'n'k'}^{s2}) - (\lambda_s^* + \lambda_t^*) P_{m'n'k'}^{s2}
\]

\[
\frac{w_{k'}}{2} \log(1 + a_{m'} P_{m'n'k'}^{s1}) + (\lambda_s^* + \lambda_t^*) P_{m'n'k'}^{s1} \geq \frac{w_{k'}}{2} \log(1 + c_{m'k'} P_{m'n'k'}^{s2})
\]

\[
\frac{w_{k'}}{2} \log(1 + a_{m'} \min\{P_s, P_l\}) + \frac{w_{k'}}{\alpha} \geq \frac{w_{k'}}{2} \log(1 + c_{m'k'} P_{m'n'k'}^{s2}) - 4 + 4a_{m'} \min\{P_s, P_l\} \geq 1 + c_{m'k'} P_{m'n'k'}^{s2}
\]

\[
4 + 4a_{m'} \min\{P_s, P_l\} \geq \frac{w_{k'}c_{m'k'}}{\alpha(\lambda_s^* + \lambda_t^*)}.
\]

(7.26)

Hence, \( \lambda_s^* + \lambda_t^* \geq \frac{w_{k'}c_{m'k'}}{4\alpha(a_{m'} \min\{P_s, P_l\} + 1)} \), so \( \lambda^* \in \mathcal{R}_1 \).

- Else, the only scenario left is when \( c_{mk} = 0 \) for \( \{m, k : \phi^*_m(\lambda^*) = 1\} \). In this case, there exists \( b_{nk} > 0 \), since otherwise the achieved sum-rate is uniformly zero. Then for any \( (m', n', k') \) such that \( \phi^*_m(\lambda^*) = 1 \),

\[
\min\{P_s, P_l\} \geq \sum_{m, n, k} P_{mnk}^{ss}(\lambda^*) \geq P_{m'n'k'}^{ss}(\lambda^*)
\]

\[
\geq \frac{w_{k'}b_{n'k'}}{\alpha(b_{n'k'}(\lambda_s^* + \lambda_t^*) + a_{m'}(\lambda_s^* + \lambda_t^*))} - \frac{1}{a_{m'}}.
\]

(7.27)
which implies that \((\lambda^*_s + \lambda^*_t) + \frac{a_{mt}}{b_{nm}} (\lambda^*_r + \lambda^*_t) \geq \frac{w_{km}}{\alpha (\min\{P_s, P_t\} + \frac{1}{c_{mk}})}\). Considering the extreme case for the slope and intercept of this linear inequality, we have \(\lambda^* \in \mathcal{R}_2\).

### 7.2.3 Proof of Lemma 4.3

To find an upper bound for \(\|\lambda^*\|_2\), we consider the following cases for the activation pattern of the individual power constraints (4.4) and (4.5) and the total power constraint (4.6) at global optimum.

- **Neither (4.4) nor (4.5) is active:** In this case, (4.6) must be active, since otherwise there would be more power to increase the sum-rate. Thus, we have \(\lambda^*_s = \lambda^*_r = 0\) and
  \[
  P_t = \sum_{m,n,k} P_{ms}^s (\lambda^*) + P_{mr}^r (\lambda^*). 
  \]  

Since \(\phi_{mnk}^* (\lambda^*) \leq 1\), substituting \(\phi_{mnk}^* (\lambda^*) = 1\) and \(\lambda^*_s = \lambda^*_r = 0\) into (4.18), and considering all possible scenarios of (4.18), we have

\[
P_t \leq \sum_{m,n,k} \max\{\frac{w_k}{\alpha \lambda^*_t} - \frac{1}{a_m}, \frac{w_k}{\alpha \lambda^*_s} - \frac{1}{c_{mk}}, \frac{w_k}{\alpha \lambda^*_t}\},
\]

\[
= \sum_{m,n,k} \frac{w_k}{\alpha \lambda^*_t} = \frac{N^2}{\alpha \lambda^*_t}. 
\]  

Hence, we have \(\lambda^*_t \leq \frac{N^2}{\alpha P_t}\), so that

\[
\|\lambda^*\|_2 \leq \frac{N^2}{\alpha P_t}. 
\]  

- **Both (4.4) and (4.5) are active:** We have
  \[
P_s = \sum_{m,n,k} P_{mnk}^{ss} (\lambda^*),
  \]
  \[
P_r = \sum_{m,n,k} P_{mnk}^{rs} (\lambda^*). 
\]

Again, substituting \(\phi_{mnk}^* (\lambda^*) = 1\) into (4.18), and considering all possible scenarios, we conclude that

\[
P_s \leq \sum_{m,n,k} \max\{\frac{w_k}{\alpha (\lambda^*_s + \lambda^*_r)} - \frac{1}{a_m}, \frac{w_k}{\alpha (\lambda^*_s + \lambda^*_r)} - \frac{1}{c_{mk}}\},
\]
that subgradient updating is performed over channel gains so that the weighted rate function.

We first note that there must exist one path \((7.2.4)\) Proof of Lemma 4.4

Summarizing the three cases above, we have

\[
\left\{ \sum_{m,n,k} \frac{w_k}{\alpha(a_m - c_{mk} + (a_m - c_{mk})(\lambda^*_s + \lambda^*_r))} \right\} \leq 0
\]

\[
\left\{ \sum_{m,n,k} \frac{w_k}{\alpha(a_m - c_{mk} + (a_m - c_{mk})(\lambda^*_s + \lambda^*_r))} \right\} \leq 0
\]

\[
\left\{ \sum_{m,n,k} \frac{w_k}{\alpha(a_m - c_{mk} + (a_m - c_{mk})(\lambda^*_s + \lambda^*_r))} \right\} \leq 0
\]

Hence, we have \(\lambda^*_s + \lambda^*_r \leq \frac{N^2}{\alpha P_s}\), and \(\lambda^*_s + \lambda^*_r \leq \frac{N^2}{\alpha P_r}\), so that

\[
\|\lambda^*\|_2 \leq \frac{\sqrt{2}N^2}{\alpha \min\{P_s, P_r\}}.
\]

- **Only one of (4.4) and (4.5) is active:** We have either \(\lambda^*_s = 0\) and (7.32), or \(\lambda^*_s = 0\) and (7.33). For both cases, an upperbound for \(\|\lambda^*\|_2\) is given by (7.34).

Summarizing the three cases above, we have

\[
\|\lambda^*\|_2 \leq \frac{\sqrt{2}N^2}{\alpha \min\{P_s, P_r, P_l\}}.
\]

7.2.4 Proof of Lemma 4.4

We first note that there must exist one path \((m, n, k)\) with non-degenerate user weight and channel gains so that the weighted rate function \(w_k R(m, n, k)\) is not uniformly zero, i.e., \(w_k > 0, a_m > 0, b_{nk} + c_{mk} > 0\). In Algorithm 1, subgradient updating is performed over two cases, either there exists some \(c_{mk} > 0\) and \(\lambda^{(l)} \in \mathcal{R}_1\), or \(c_{mk} = 0\) for all \(m\) and \(k\) and \(\lambda^{(l)} \in \mathcal{R}_2\).

In the former case, let \(\epsilon_1 = \frac{\min_{\{k: a_m > 0\}} w_k \min_{\{m: a_m > 0\}} a_m \min_{\{m, k: c_{mk} > 0\}} c_{mk}}{4\alpha \min_{\{m: a_m \min\{P_s, P_l\} + 1\}} a_m \min_{\{m, k: c_{mk} > 0\}} c_{mk}}\). The projected subgradient updating is performed over \(\lambda^{(l)}_s + \lambda^{(l)}_r \geq \epsilon_1\). Since there exists some \(m\) and \(k\) such that \(w_k > 0, a_m > 0, \) and \(c_{mk} > 0\), we have \(\epsilon_1 > 0\). From (4.18), we see that in all scenarios

\[
P_{m_{nk}}^{ss}(\lambda^{(l)}) \leq \frac{w_k}{\alpha \lambda^{(l)}_s + \lambda^{(l)}_r} \leq \frac{w_k}{\alpha \epsilon_1} < \infty,
\]

\[
P_{m_{nk}}^{ss}(\lambda^{(l)}) \leq \frac{a_m - c_{mk}}{b_{nk}} P_{m_{nk}}^{ss}(\lambda^{(l)}) \leq \frac{w_k(a_m - c_{mk})}{\alpha \epsilon_1 b_{nk}} < \infty.
\]
The second inequality above hold since \( P_{mnk}^{rs}(\lambda^{(l)}) = 0 \) when \( b_{nk} = 0 \).

In the latter case, let \( \epsilon_2 = \frac{\min_{(k:w_k>0)} w_k}{\alpha \left( \min \{ P_s, P_t \} + \min_{\{m:a_m>0\}} a_m \right)} \). The projected subgradient updating is performed over \( (\lambda_s + \lambda_t) + \frac{\min_{a_m>0} a_m}{\max_{n,k} b_{nk}} (\lambda_r + \lambda_t) \geq \epsilon_2 \). Since there exists some \( m \) and \( k \) such that \( w_k > 0 \), \( a_m > 0 \), we have \( \epsilon_2 \). Furthermore, \( b_{nk} > 0 \) since From (4.18), we see that in all scenarios

\[
\begin{align*}
P_{mnk}^{ss}(\lambda^{(l)}) & \leq \frac{w_k}{\alpha (\lambda_s^{(l)} + \frac{a_m}{b_{nk}} \lambda_t^{(l)})} \leq \frac{w_k}{\alpha \epsilon_2} < \infty, \\
P_{mnk}^{rs}(\lambda^{(l)}) & \leq \frac{a_m - c_{mk}}{b_{nk}} P_{mnk}^{ss}(\lambda^{(l)}) \leq \frac{w_k (a_m - c_{mk})}{\alpha \gamma b_{nk}} < \infty.
\end{align*}
\]

The second inequality above hold since \( P_{mnk}^{rs}(\lambda^{(l)}) = 0 \) when \( b_{nk} = 0 \).

We note that the bounds in (7.36)-(7.39) are not functions of \( m \) and \( n \). Substituting the bounds of either of these two cases into (4.31), and noting that only \( N^2 \) paths are chosen, we have \( \| \theta(\lambda) \|_2 = O(N^2) \).

### 7.3 Appendix C

#### 7.3.1 Proof of Lemma 5.1

Consider \( P^*_t(k) > 0 \) for any \( k > 0 \). We only need to show that there exists \( P < P^*_t(k) \) such that \( B^*(P) = k - 1 \).

The required \( P \) can be found using the following procedure. Within the solution that renders \( P^*_t(k) \), pick any path \( (m,n) \) such that \( \phi^*_mn = 1 \) and \( b^*_mn > 0 \), which implies \( p^*_mn > 0 \). From (5.4) and the concavity of \( R(\cdot) \), we see that there exists \( p_{mn} < p^*_mn \) such that the corresponding data rate \( b_{mn} = b^*_mn - 1 \). Let \( P = P^*_t(k) - p_{mn} + p_{mn} \). Then \( P < P^*_t(k) \), and the maximum sum rate given \( P \) is no worse than \( k - 1 \), i.e., \( B^*(P) \geq k - 1 \). However, since \( P^*_t(k) \) is the minimum required power to achieve sum rate \( k \), we also have \( B^*(P) < k \) for any \( P < P^*_t(k) \). Therefore, \( B^*(P) = k - 1 \).
By taking the minimum of all \( P \) such that \( B^*(P) = k - 1 \), we obtain \( P_t^*(k - 1) < P_t^*(k) \). With further induction, we see that \( B^* \) is a unit-step increasing staircase function of \( P_t \), with integer-valued corner points at \((P_t^*(k), k)\).

### 7.3.2 Proof of Lemma 5.2

Given any \( \lambda \geq 0 \), consider the maximization problem (5.6) with \( P_t = P_{\lambda}^*(\lambda) \). The power constraints (5.2) and (4.6) in this problem are automatically satisfied. Furthermore, the other constraints are the same as those in the maximization problem (5.10). Further comparing the objective functions, we see that maximizing the objective function of (5.10), given \( \lambda \) and \( P_{\lambda}^*(\lambda) \), is exactly the same as maximizing the objective function of (5.6). Hence, we have

\[
B^*(\lambda) = B^*(P_{\lambda}^*(\lambda)) , \tag{7.40}
\]

i.e., the point \((P_{\lambda}^*(\lambda), B_{\lambda}^*(\lambda))\) always falls on the curve \( B^*(P_t) \).

Furthermore, for all \( \lambda > 0 \), if \( B_{\lambda}^*(\lambda) = B \) for some \( B \in \mathcal{M} \), then in order to maximize the objective in (5.10), \( P_{\lambda}^*(\lambda) \) should be the minimum allowed sum power, subject to (2.4), (5.1), (5.2), (5.4), and \( \sum_{m,n} \phi_{mn}^*(\lambda) b_{mn}^*(\lambda) = B \). This is of the same form as (5.7). Hence, \( P_{\lambda}^*(\lambda) = P_t^*(B) \).
Bibliography


MIMO-OFDM relay links,” IEEE Trans. Wireless Commun., vol. 6, no. 8, pp. 2798–2802,
Aug. 2007.

design for a two-hop relay OFDM system,” in Proc. IEEE Int. Conf. Communications
(ICC), June 2009.

[10] G. Li and H. Liu, “Resource allocation for OFDMA relay networks with fairness con-

transmission in OFDMA-based cooperative relay networks,” EURASIP J. Appl. Signal
Process. (USA), 2009.

[12] Y. Cui, V. Lau, and R. Wang, “Distributive subband allocation, power and rate control for
relay-assisted OFDMA cellular system with imperfect system state knowledge,” IEEE

2010.


in selection-based OFDM networks,” in Proc. IEEE Int. Conf. Communications (ICC),
June 2011.


